An Application of Multilevel Modelling Techniques to the Longitudinal Study of Student Progress in a Modular Degree Course

PhD thesis

Vanessa Simonite

Institute of Education, University of London

For Marian and Ches

Acknowledgements

I would like to thank the School of Computing and Mathematical Sciences, Oxford Brookes University, for supporting this research through the Staff Development Fund and Professor Harvey Goldstein for his support as a supervisor.

Thank you Annie Jo, Laura, Tom and Jeremy for your support and entertainment too.

Abstract

This thesis presents a longitudinal study of undergraduate achievement within a modular first degree course, analysing the academic records of a cohort of students who graduated from the Modular Degree Programme at Oxford Brookes University. Multilevel models are fitted to the marks achieved by members of this cohort in each module taken. Level 1 units are individual module entries, nested within occasions within individual student's programmes. These models were fitted by maximum likelihood and used to study the effects of both student and module characteristics on performance. The effects of these factors on mean marks, on the consistency of students performance and on the variation between students were studied by including complex variation at level 1 and random effects at student level in the models. In addition, individual progress charts were fitted, showing how patterns of progress vary from one student to another.

Reviewing the hierarchical structure, it was found that a more complex, cross-classified structure is needed to represent the data accurately. This recognises that individual module entries are clustered within modules, as well as within students. Fitting large multilevel cross-classified models is computationally difficult, however newly developed MCMC estimation techniques allowed a model based on the more complex structure and including random effects and complex variation to be fitted. This analysis shows how MCMC estimation techniques can be used to fit a large cross-classified multilevel model, incorporating random effects and complex variation. The results obtained describe students' progress over the period of their degree course and measure the effects, other things being equal, of factors such as

assessment methods, age and subject on mean levels of achievement, consistency of performance and the variation between students, providing a model for future studies of achievement within a modular framework.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Aims

Modular degree programmes are now the most common form of undergraduate education within the UK (HEQC, 1996). Modular courses share a common framework, in which students follow programmes of study consisting of modules chosen in accordance with the regulations governing their chosen degree. As students pursue their programmes of study, assessment takes place within individual modules. The results of these assessments provide each student with regular feedback on their performance, determine a student's status on the course and, once sufficient 'credits' have been accumulated, are used to classify their degree. This thesis will examine the factors affecting the academic achievement of students within modular first degree programmes by analysing the detailed records of achievement for a cohort of students who graduated from the Modular Degree Programme at Oxford Brookes University. Within modular programmes, students undergo multiple assessments on each of a number of occasions and the data recording the outcomes of these assessments is highly structured. By analysing the results achieved in each module contributing to a student's record and by taking into account the structure of this data, the first aim of this thesis is to provide a model for future studies of student achievement within modular programmes.

An understanding of the relationships between achievement and a student's background is important in considering whether equal opportunities are available to undergraduates on a modular degree programme and in targeting learning support at the students who need it most. As characteristics such as age, gender and entry qualifications may not be independently distributed, controlling for the effects of other variables is important when comparing the achievements of groups of students. Similarly, as students' programmes vary, comparisons between groups of students need to take into account the effects of any differences in their programmes. The research literature on undergraduate achievement shows that while several performance-related factors have been identified, the analyses on which these findings are based generally fail to control for the effects of factors related to those being studied. The second aim of the research presented here is to re-evaluate the effects on performance of variables studied in earlier research after controlling for more concomitant variables.

As modular degree courses assess students on several occasions, changes in their performance during the course of their first degree can be studied. The third aim of the research presented here is to model student progress and to determine how individual student's patterns of progress vary.

Further aims are concerned with variations in performance: between students and within student records, between the marks awarded in different modules. The fourth aim of this thesis is to determine which student and module characteristics influence the variation between students and to measure their effects. The fifth aim of this thesis is to examine factors affecting the consistency of individual students' performance.

1.2 Summary of Thesis

Chapter 2 provides a review of the research literature on undergraduate achievement, identifying the need for more research. Previous research findings are based on analyses that generally fail to control for the effects of factors related to those being studied or use measures of achievement which discriminate poorly between students. The use of cross-sectional data in earlier studies of achievement in higher education means that there is no empirical evidence of how consistently students perform or of whether some groups of students perform more consistently than others. This consistency is an important feature of achievement within a modular course: a student's continued enrolment and their graduation both depend on their success in accumulating module credits and this in turn will depend on the level and consistency of their performance. Chapter 2 also includes a discussion of the properties of degree classification systems, concluding that degree classifications partly depend on the variation in the marks achieved within an individual's programme, so factors that influence this variation will also influence graduation and final degree awards.

The research presented here is based on analyses of data extracted from the student record system at Oxford Brookes University and consisting of the student records, including the marks achieved individual modules, of a sample of 496 students who graduated from the Modular Degree Programme in July 1997. Chapter 3 provides a detailed explanation of the structure of the Modular Degree Programme at Oxford Brookes University and describes the selection of the sample of graduates whose records are analysed here. Further sections in chapter 3 summarise the characteristics of the cohort of graduates and their programmes of study.

Chapter 4 proposes models in which the structure of the data is represented by a three level hierarchical model, in which marks for each module entry are clustered within terms within the student's programme. The model includes parameters representing the effects of student and module characteristics in a variety of ways. To achieve the aims of the thesis, student and module characteristics are represented as having fixed and random effects; individual patterns of progress and complex variance structures at student level and at the level of individual module entries (level 1) are also included in the model. The model proposed was fitted to the data in stages, adding different kinds of effects and reviewing the contributions of each variable to the model at each stage. Chapter 4 presents the results obtained at each stage and examines the residuals associated with the final selection of variables and effects.

Using a model based on a multilevel structure allows the relationships between the individual marks awarded to the same individual, at the same or different points within their programme, to be recognised. The importance of recognising such relationships is well known (Goldstein, 1995), but there are other relationships between the marks awarded to this sample of graduates, since in some cases, the marks were awarded within the same modules. Chapter 5 reviews the structure of the data and concludes that it is more complicated than the simple hierarchy proposed in chapter 4. Models based on a multilevel, cross-classified structure are proposed but the more complex structure greatly increases the computational demands of fitting the model and, as a result, alternative estimation methods have to be considered. MCMC estimation techniques are capable of fitting models based on the more complex structures to the full dataset and were applied here, using software under development (Browne, 2000). The cross-classified model, including variables and effects identified as important in the hierarchical analyses, and fitted to the student record data using MCMC estimation techniques, is presented in chapter 6. Chapter 7 evaluates the findings of the final analysis, showing how it enables the first four aims of the thesis to be met. The methodological approach used to obtain the research findings demonstrates the range and complexity of the questions that can be tackled using student record data and in doing so provides a model for future analyses of student achievement within modular programmes.

Chapter 2

Review of Research Literature on the Academic Achievement of Undergraduates

2.1 Introduction

This chapter is concerned with the methods and findings of previous research studying the academic achievement of undergraduates and with two key issues related to modular degree courses: comparability and degree classification systems. Section 2.2 of this chapter reviews the methods used to investigate variations in academic achievement in first degree courses and finds that these studies have often been constrained by the data available from official or institutional statistics and sometimes by the methods of analysis chosen. The findings of these studies are reviewed in section 2.3, which identifies some of the limitations and gaps in the evidence on which our understanding of how students perform in higher education is based.

Section 2.4 describes the growth in modular degree courses in the United Kingdom. Modular schemes involve an implicit assumption of comparability between modules: this issue is discussed in section 2.4.1. The aggregation of module marks or grades to decide a student's degree classification is another important feature of a modular degree programme: the implications of different classification systems are discussed in section 2.4.2.

Finally, section 2.5 identifies the gaps in current knowledge, which will be addressed by the research presented here.

2.2 Methods used to study the academic achievement of undergraduates

Researchers have used primary and secondary data to study academic achievement in first degree courses and the student, course and institutional characteristics that may influence it. Official statistics provide an annual census of students in higher education, allowing researchers to look at patterns of variation across the whole of the higher education sector and the records maintained by individual institutions enable case studies to be carried out. Studies based on secondary analyses of official statistics or institutional records avoid labour intensive fieldwork and problems of non-response but there are other difficulties, such as the restricted set of variables available. Researchers who collect primary data are free to specify the variables to be collected but may find it hard to achieve a high response rate or to obtain accurate responses when collecting information retrospectively. Where studies are carried out within an individual course or module the teaching timetable can be used to contact and administer questionnaires to students but this approach is not available or appropriate in all cases.

2.2.1 Measuring academic achievement

Existing studies of academic achievement in first degree courses do not appear to have constructed their own measures of academic attainment, and instead use the results of assessments carried out within students' degree courses. Researchers analysing official statistics and frequently those using institutional data use degree class, or a function of it, to measure academic achievement. Although degree class is important, since it is used in selection for employment and for postgraduate education, Winter (1993) and Elton (1998) anticipate the replacement of degree classifications by a pass/fail system accompanied by profiling. Degree class does not discriminate well between students; graduates are allocated between four categories: first class, upper second lower second, third and pass/ordinary. The low level of discrimination between students is illustrated by figures reported by the Higher Education Statistics Agency (HESA, 1997), showing that 41% of students who graduated in the UK in 1995/6 were awarded upper second class degrees and 36% were awarded lower second class degrees. In some analyses, degree classes are combined to provide a binary measure, which distinguishes between students who achieve a 'good' degree (that is, a first class or upper second class degree) and those who do not. A further disadvantage is that it is not clear what degree class precisely measures and this will, in any case, vary from one degree course to another. The reliability of degree classifications does not appear to have been estimated.

Analyses of official statistics may involve the degree results for students graduating in different years, in different subjects, from different institutions, or combinations of more than one of these. For example: Bourner and Hamed (1987), analysing the degree results of students awarded degrees in 1983 by CNAA, combined results from a wide range of institutions and courses to study the relationship between students' ages and degree class and Tomlinson and MacFarlane (1995) distinguished only between 'Arts' and 'Science' degrees in a study of degrees awarded by 'old' universities. Analyses such as these assume first, that meaningful comparisons can be made between the achievements of students following different degree courses, however different the syllabi, methods of assessment or institutional cultures, and second, that the boundaries between degree classes define equivalent levels of attainment in all the courses concerned. The question of comparability, which is crucial to the interpretation of the findings of previous research and the analyses reported later, is discussed more fully later in this chapter.

Binary and ordinal measures based on degree class require more complex statistical analysis than dependent variables measured on a continuous scale. Some researchers allocate numeric scores to different degree classes in order to 'transform' degree class to a continuous variable (for example Hoskins et al, 1997). The difficulty here is that the values associated with each degree class are arbitrarily chosen, so that a different set of values will produce different estimates of effects on mean performance, different standard errors and therefore different conclusions. Consequently, analyses using this approach are extremely difficult to interpret.

In some studies, researchers analyse students' marks or grades for a component of their degree (e.g. for the first year of a course or for a single module). These measures of performance are more successful than degree class in discriminating between students, but inevitably the scope of the research is restricted, so that for example, the large samples needed for multi-factor analyses are not available and similarly, comparisons between institutions are ruled out. The reliability of these assessments is generally unknown and, if the outcomes from different assessments are analysed together, then the question of comparability between components or modules needs to be considered.

2.2.2 Absence of Key Variables

The variables available for secondary analyses of institutional data or official statistics are those collected by the relevant administrative system. Apart from the degree course and institution attended, these are likely to consist of standard demographic variables such as age, sex, parents' social class, domicile and ethnic origin combined with the students' qualifications on entry to the course and the type of previous educational establishment attended. Thus no information is available describing the student's circumstances or behaviour during their degree course, official statistics also lack information describing the teaching and assessment regimes experienced by the student.

The absence of potentially important explanatory variables in analyses of degree outcomes (or other measures of students' performance) means that, since these are observational studies, variations in performance cannot be attributed unambiguously to known factors since they may be produced by the influence of unobserved variables. This problem is accentuated in studies such as Bourner and Hamed (1987) or Woodley (1984) where the effects of only one or two potentially influential factors on performance are considered at a time, even though, in both cases, additional variables were available.

Where differences in the performance of sub-groups of students are known to occur, possible explanations can be suggested or supported by empirical data showing other ways in which the sub-groups differ. Our understanding of factors influencing undergraduate performance is limited by the tendency for studies that are not directly concerned with academic achievement to be concerned with only one aspect of students' experiences. Studies of undergraduates that do not compare sub-groups cannot contribute this kind of evidence. For example, Ford et al (1995) studied students' part-time paid work and reported the occupations, hours and rates of pay for undergraduates sampled at four universities in the UK. A negative association between part-time work and student's grades was found by Paton-Salzburg and Lindsay (1996), but since Ford et al (1995) made no comparisons between men and women or between mature and other students, we cannot use their findings to assess how different patterns of part-time employment might lead to differences in the academic achievements of male and female or traditional and mature students.

2.2.3 Sample Selection and Response Rates

The selection of cases for study determines the population to which conclusions refer. Where researchers use official statistics covering a range of institutions, their conclusions may apply very broadly, subject to the assumption that degrees awarded by different institutions are comparable in standard.

Whatever the method of data collection, when degree class is used to measure academic achievement, the students who contribute data are a biased selection from the cohort entering higher education at the same time, since those who do not graduate are automatically excluded. Other biases can occur as the result of non-response, particularly in surveys of students where non-response rates are often high. A typical example is the study carried out by Ford et al (1995), a postal survey of undergraduates, which obtained an overall response rate of 60%. Studies in which data are collected within a lecture might be expected to achieve a higher response rate, given a 'captive' audience, but this is not always the case; for example, Richardson (1993) reports response rates of 64% and 54% when collecting information from first year students during lectures on two occasions. These response rates compare unfavourably with the census style coverage produced by agencies such as HESA or the USR.

Non-response bias may occur in official statistics and institutional data as the result of item non-response; for example, Rudd (1987) found that students for whom parents' social class had not been measured were not randomly distributed across classes, and therefore students for whom this variable was recorded were a biased sample of the student population as a whole.

2.2.4 Level of Analysis

Analyses of official statistics are based on the degree results for large numbers of students, graduating from many institutions. Although the education system within which degrees are awarded has a clear hierarchical structure, with students learning within institutions, analyses modelling this structure have yet to be published. Hence the research literature on the relationship between degree results and entry qualifications includes some studies in which the student is the unit of analysis (for example, Bourner and Hamed, 1987) and others in which the institution is the unit of analysis (for example, Bee and Dolton, 1985 or Johnes and Taylor, 1987) but none in which a multilevel approach has been used. As a result, the extent to which relationships between degree results and other factors vary between institutions is unknown, since analyses at student level do not allow for such variations, while analyses of institution level data generate estimates which are unreliable, having large standard errors (Goldstein, 1995). A further problem is that relationships between variables at institution level may bear no relation to the relationships between the same variables at student level.

2.2.5 Changes in Higher Education

Changes in higher education in the UK during the last 30 years include increases in student numbers, increases in the proportion of women and mature students, the creation of the 'new' universities, changes in assessment practices and the widespread development of modular courses. On the one hand these changes generate interest in the research literature on related topics, but on the other hand the same changes raise questions about whether earlier research findings remain valid. For example, an increase in the proportion of female students taking engineering degrees may lead to interest in gender differences in achievement in such courses and research studies carried out earlier may provide estimates of the gender difference in achievement on engineering courses. The problem is that the recruitment of increased numbers of women may have been achieved by drawing students from a wider population than before and will change the mixture of male and female students on engineering courses. These changes may lead to changes in the differential in performance between male and female students, so that the findings of earlier studies no longer apply. Research findings are 'of their own time' and extrapolating from one period in higher education to another requires caution.

Sometimes the time that has elapsed between assessments and the analysis of their outcomes means that the potential for investigating a particular topic is limited. When a new admission policy is adopted, at least three years must elapse before the first cohort of students to be admitted under the new admission policy graduates, enabling the effects of the new policy on degree outcomes to be evaluated. In practice, even longer periods of time may elapse between a cohort's graduation and the publication of research based on their achievements; for example, Johnes' study, published in 1992 is based on a sample of individuals who graduated in 1980, while

Woodley's study of the achievement of mature students, published in 1984, is based on the degree results of students who graduated in the 1970s.

Changes in higher education, for example in admission procedures, the student population or methods of assessment, can lead to worries about 'standards' and whether they may be declining so that the educational value of a degree compares poorly with the value of similar degrees in the past. The next section discusses this and other aspects of the comparability between degree awards.

2.2.6 Assumptions of Comparability between degree awards

Several of the studies whose findings are discussed in section 2.3 analyse the degree results of students graduating in different years, from different institutions, in different subjects, or combinations of these three. In studies that combine data for students who have been assessed differently, or in relation to different syllabuses, the conclusions are based on the assumption that these assessments are comparable. This section discusses what is meant by comparability of standards: between degrees in the same subject awarded by different institutions, between the same degrees awarded in different years and between degrees awarded in different subjects. Johnson (1988) notes that in higher education, there have been few attempts to investigate comparability between degree awards, and contrasts this with the large body of work on comparability between public examination results in secondary education. In spite of the differences between the grading of public examinations by examination boards and the awarding of degrees by different institutions, much of the work on public examinations can usefully be applied in the context of higher education.

2.2.6.1 Comparability between degrees awarded by different institutions

Comparability between degree awards in different institutions, is one of the central purposes of the external examining system, as the CVCP (1984) Code of Practice for external examiners states:

"*The purposes of the external examiner system are to ensure, first and most important, that degrees awarded in similar subjects are comparable in standard in different universities in the United Kingdom, though their content does of course vary; and, secondly, that the assessment system is fair and is fairly operated in the classification of students.*"

The requirement for comparability between institutions in 'similar' subjects and the reference to variations in course content suggests that meaningful comparisons can be made between courses whose syllabuses may differ substantially. This optimism is not shared by those who have studied comparability in relation to secondary education (Nuttall, 1979;Goldstein and Cresswell, 1996;Newton, 1997). Even different emphases in either questions or the marking schemes used in examinations covering the same syllabus, can prevent meaningful comparisons between students assessed by different examinations (Johnson, 1988). Similarly, it is difficult for examiners to identify comparable attainments on courses that use different styles of teaching or assessment. Identifying comparable standards of achievement in courses in 'similar' subjects or where the course content varies is therefore problematic. For now, comparability will be considered for degree awards in the same subject, assuming that the syllabuses are sufficiently similar for meaningful comparisons to be made and the question of wider comparisons will be considered later.

One definition of comparability is that 'degrees awarded by different institutions are comparable if they produce the same distributions by degree class'. (Cresswell's (1996) 'no-nonsense' definition). If this 'norming' definition is used it becomes impossible to monitor differences between institutions since they will automatically produce the same distributions of results; also selective admission to first degree courses means that institutions do not draw their students from the same population and therefore it is unreasonable to require them to produce the same distribution of awards. A further criticism of this approach is that it makes no reference to the content and educational value of either the syllabuses or students' assessed work (Goldstein and Cresswell, 1996).

Since student intake varies between institutions, a better definition of comparability might be that 'degree awards in the same subject, given by different institutions are comparable if the distributions by degree class are the same in both institutions for students with the same entry qualifications'. 'Value-added' analyses, in which within-subject comparisons between institutions are adjusted for student intake, have been carried out using data aggregated to institutional level (Chapman, 1996; HEQC, 1996a). Here the differences between institutions were measured by comparing the residuals for the institutions, found by adjusting the percentage of good degrees to take into account awarded the differences between institutions in the entry qualifications of their students. One problem with this approach is that the differences between institutions in their residuals could be explained by institutional or student characteristics not included in the model rather than their using different standards. This will be a problem even if results are adjusted for the effects of additional variables. The HEQC (1996a) report is suitably cautious about interpreting the residuals, although the effects of student variables are not mentioned. A further problem is that the relationship between entry qualifications (and any other variables included) and degree results may vary from one institution to another so that the differences between institutions are not fixed, but dependent on the type of student considered. 'Differential comparability' is a fundamental problem of statistical definitions of comparability (Newton, 1997; Goldstein & Cresswell, 1996). Another general problem with statistical definitions is that they do not take into account the educational content of courses or assessments or the way that examiners' responds to assessed work (Goldstein & Cresswell, 1996).

The 'value added' approach is an example of using a 'reference test', in this case entry qualifications, to define comparability. With this definition, two assessments are considered comparable if students who take one assessment achieve the same outcome as students with the same score for the 'reference test' achieve on the other. Underlying this approach is the assumption that the assessments being compared are 'uni-dimensional', that is, that they measure the same underlying attribute (Goldstein and Cresswell, 1996) In the case of degrees awarded by different institutions, where students are assessed in relation to different syllabuses, this assumption is very unlikely to be true.

2.2.6.2 Comparability between degrees awarded in different years

Maintaining standards over time is important for individuals as well as institutions since it is common for individuals who graduated in different years to compete with each other for jobs or places on postgraduate courses. For selection based on degree awards to be fair, there must be comparability of standards between degrees awarded in different years. A report for the HEQC (1996a) expresses the view that changes in course content, assessment methods and discoveries and developments within disciplines prevent meaningful comparisons of standards over time. For year to year and short-term comparisons, courses may be sufficiently similar for meaningful comparisons to be made.

Over time there have been changes in the distribution of degrees awarded by class; the proportion of good degrees has increased and in many subjects the modal degree class has shifted from a lower second class degree to an upper second class degree (see for example, HEQC, 1996a). The direction of this change is surprising in the context of increasing student numbers and decreasing per capita funding and has therefore lead to questions about changes in standards (MacFarlane, 1992; Gibbs & Lucas, 1997; Chapman, 1998; NCIHE, 1997 (Dearing report)). The upward drift in degree class has been attributed to the increasing use of coursework assessment (Gibbs & Lucas, 1997), interpreted by Elton (1998) as showing the incompetence of examiners who should have compensated for the effects of changes in assessment instruments and by Chapman (see HEQC, 1996a) as showing that continuous assessment is fairer to students. These varied conclusions illustrate the problems of using empirical data alone to draw conclusions about standards; the patterns of results obtained stimulate observers to produce competing explanations but do not help them to choose between them.

A 'norming' approach to defining comparability over time, awarding the same proportions of degrees in each class every year, makes it impossible to monitor the effects of changes in institutional, course or student characteristics related to performance. Changes in the student population can be dealt with by defining awards as comparable if they produce the same distribution of outcomes for students with the same entry qualifications. This is the reference test approach, but this requires the assumption that the assessment leading to the award of a degree is uni-dimensional, an assumption that is not true in this context. There is also the problem of differential comparability; changes in results from one year to another may imply that different standards were used to assess students in different years or that there were changes in the population from which students are drawn. This means that with no educational assessment of the content of the course or the students assessed work, equivalent standards cannot be defined.

2.2.6.3 Between subject comparability

This aspect of comparability is the one that has given the most trouble (Nuttall, 1979; Johnson, 1998; Goldstein & Cresswell, 1996). Assessments in different subjects are concerned with qualitatively different attainments, so that identifying quantitatively equivalent levels on each is, logically speaking, impossible. There appears to be no support in higher education for requiring degrees in different subjects to use comparable standards (Silver et al, 1995; HEQC, 1996b), however, widespread modularisation of degree courses has lead to increased concerns about differences in marking practices between subjects at module level (HEQC, 1996); and these will be discussed in section 2.3.6.

2.2.6.4 A subjective definition of comparability

The last sub-sections show that statistical techniques for defining comparable standards for degree awards, between institutions, over time or between subjects, lead to problems and inconsistencies. This agrees with the conclusions reached in the work on comparability in public examinations in secondary education, where examination boards in the 1970s, having been unable to find a suitable statistical technique, adopted procedures based on expert judgement. A subjective definition of comparability is compatible with the external examiner system in higher education, which is based on an implicit assumption that comparability is a matter of academic judgement.

Cresswell (1996) gives a formal definition of comparability in terms of expert judgement:

> *'Two examinations have comparable standards if candidates for one of them receive the same grades as candidates for the other whose assessed attainments are accorded equal value by awarders accepted as competent to make such judgements by all interested certificated users.'*

The definition involves three key elements: the students' attainments, the awarders, whose role is evaluate, or attach values to students attainments and the interested users who have the power to accept or reject the awarders' competence.

In public examinations, the awarders comprise a chief examiner, who sets the examination paper and a panel of subject experts. For degree awards, the awarders could be defined as the internal and external examiners and other members of examination boards or committees with responsibility for deciding or confirming the degree awards. Defining the awarders in this way recognises the role of the course team (represented by internal examiners and other members of the examination committee) as well as the external examiner, in setting and maintaining standards.

The idea that standards are negotiated between representatives of the institution and of the academic community as a whole is also compatible with views expressed in the Silver report (Silver et al, 1995, p.41) and the Dearing report (NCIHE, 1997).

The 'interested users' for degree awards include students and staff in the awarding and other institutions, employers, admission tutors for postgraduate courses, professional associations and other interested groups. An important feature of Cresswell's (1996) definition is that comparability is defined in a social context and is not achieved unless the users accept the awarders as competent and comparability may be achieved for some users but not others.

In public examinations, the work which awarders carry out in order to reach their decisions involves comparing course syllabuses, evaluating the contents of assessment instruments, considering samples of marked scripts and reviewing relevant statistical data. In Cresswell's (1996) view, this work is of paramount importance and has two functions: first to ensure that the awarders' decisions, are reliable, in the sense that other awarders would usually reach the same conclusions if presented with the same evidence and second, to provide evidence of the awarders' competence in order to maintain their acceptance by interested user groups. Similar activities have traditionally been carried out by external examiners but there is now much greater diversity - in students, courses and institutions than in the higher education sector for which the external examiner system was designed (HEQC, 1996b; Silver et al, 1995). To be able to ensure comparability between institutions, the external examiners must be able to refer to, and therefore know about, implied common standards used across all institutions offering degrees in their subject. Johnson (1988) questions how any individual could acquire such knowledge and therefore proposes that individual external examiners could be replaced by teams who, collectively, would have wider knowledge. Within modular degree programmes, the module level focus is a handicap since although modular frameworks have common features, details such as module size, structure of the academic year, number of levels, grading systems, methods of aggregation and 'packaging' of degree contents into smaller units are determined at institutional level. Evaluating attainments within different structures must be extremely difficult, particularly given the external examiner's loss of an 'overview' of students' whole programme of studies. These represent real difficulties for external examiners and also for course teams setting and maintaining standards, even where comparability is sought across a group of similar institutions rather than the whole sector.

The Silver report found support for defining comparability within subjects within groups of similar institutions rather than across all institutions (Silver et al, 1995). This can be interpreted as groups of interested users specifying the conditions under which they would be prepared to accept awarders as competent to ensure comparability. Support for the training of external examiners (Silver et al, 1995) can also be interpreted as showing demand from interested users for evidence of greater competence of external examiners. Cresswell (1996) states that comparability can be achieved, if the interested parties agree to abide by the awarders' decision, even if they do not agree with the decision itself. An example of this is that although there is widespread evidence that comparability is not established across all institutions, degrees from all institutions are accepted as equal by research councils funding postgraduate study (Silver et al, 1995).

2.2.6.5 Interpreting analyses of degree awards

With a subjective approach to defining comparability, research into achievement in higher education can be viewed as a social context, in which comparability is defined. In each study, the conclusions may be extrapolated to other contexts in which similar assumptions of comparability between degree awards are made. Hence the findings of analyses in which data are aggregated across the whole sector, will apply to other situations in which degrees awarded by different institutions are accepted as having equal value. Similarly, where the outcomes of degrees awarded in different years or in different subjects have been aggregated in order to investigate, say, the differences in performance between sub-groups of students or the effect of an interaction between age and gender, the findings can only be applied to other contexts in which these awards are assumed to be comparable. One problem in assessing the comparability of degree awards is that only relatively simple tables comparing institutions are provided by official statistics.

The use of single level techniques of analysis in the research literature means that researchers analysing data at student level have ignored the clustering of students within institutions. This involves sweeping assumptions of comparability between degrees awarded by different institutions and at the same time, provides no empirical evidence to inform a serious consideration of the standards used by different institutions.

Having discussed some of the methods and assumptions involved in research into academic achievement in higher education, the next section discusses the findings that have been produced in studies using these methods and assumptions.

2.3 Findings of studies of the academic achievement of undergraduates

2.3.1 Sex differences in academic achievement

Rudd (1984) made comparisons between the degree results of men and women who graduated from British universities in the years 1967,1978 and 1979 and found that the differences between men and women in the percentages awarded "good", that is, first class or upper second class degrees, varied from one subject to another: in arts, literature and languages, men achieved a higher percentage of 'good' degrees than women, but in other subjects, such as education, the reverse was true. Comparing men and women within subject groups, class by class, Rudd found a consistent tendency for women to obtain proportionately fewer firsts and thirds than male students studying the same subject. This pattern has been confirmed in analyses of degrees awarded by 'old' universities (Clarke, 1988; Tomlinson and Macfarlane, 1995), in studies of degree results in individual subjects (Cohen and Fraser, 1992;

Chapman, 1996) and can also be seen in recent results for all universities in the UK for students graduating in 1996/7 (HESA, 1998), in the same year as the cohort studied here. A similar pattern, of greater variation in performance amongst boys, has been observed in research carried out into achievement in schools (Gipps and Murphy, 1994).

Comparing the entry qualifications of men and women, Rudd found no explanation for the differences in their degree results and produced a number of speculative explanations for the differences between men and women, ranging from marker bias to menstruation, finally suggesting that the differences in the distribution of men and women by degree class were the result of innate sex differences in intelligence. Clarke (1988) strongly resisted Rudd's argument in favour of explanations involving marker bias and social pressures, varying between subjects and with differential effects on men and women. Clarke (1988) interpreted the superior performance of mature female students as supporting his argument, that mature women were able to resist social pressure to conform to stereotypes.

Rudd (1984), Clarke (1988) and Tomlinson and Macfarlane (1995) explored sex differences in a relatively unsophisticated way, controlling for pre-existing differences between the sexes simply by comparing their results within subject groups or within groups of students with similar A-level scores. This approach provides only limited scope for exploring differences between groups. Other researchers (Peers, 1994; Johnes, 1992;Hoskins et al, 1997; Hartley et al, 1997) have used multiple regression to investigate the effects of several potential explanatory variables, some treating degree class, their dependent variable as ordinal, others assigning scores to each class and treating the result as a continuous variable. Hoskins et al's (1997) study found statistically significant differences between men and women students in the form of a three-way interaction between age, gender and type of entry qualification (categorised as A levels/other). Hartley's (1997) study found that women performed better than men on average, with a statistically significant sex by subject interaction showing the difference to be greatest for science degrees. Both Hoskins et al's (1997) study of students at the University of Plymouth, and Hartley et al 's (1997) study of students at the University of Keele use a simple scoring system to transform degree class into a 'continuous' variable which was then analysed using parametric techniques designed to be applied to ratio level measurements. The problems associated with this treatment of degree class were discussed earlier in section 2.2.1.

Peers (1994) and Johnes (1992), treating degree class as ordinal, found only weak evidence of sex differences in degree results, other things being equal. Both authors analysed subjects or subject groups separately, avoiding the assumption of comparability between subjects. Johnes (1992) analysed data from the National Survey of Graduates and Diplomates, a sample survey of individuals who graduated from universities, polytechnics and colleges of education in 1980. Johnes used a logit model to regress degree class on a collection of variables, and carried out separate analyses for each of 13 subjects. After allowing for type of institution, type of school, sponsorship, employment between secondary and higher education, age and A-level points, the sex of the student explained significant variation in degree results in only one analysis, for language degrees. Peers' (1994) study was based on information obtained from students at the University of Manchester, divided into four subject groups. Peers (1994) found significant sex differences only in social science degrees. Peers' (1994) and Johnes' (1992) studies represent the best attempts to measure differences in the achievements of men and women graduates, after controlling for other factors, but are limited by an outcome measure which is relatively insensitive and a restricted set of explanatory variables.

The difference between the percentages of men and women achieving first class degrees is particularly marked in Oxford and Cambridge Universities; Gipps and Murphy (1994) discuss a series of papers on this subject, listing factors which have been proposed as explanations for the lower percentage of firsts achieved by women. These include the predominantly masculine culture, the lack of women academics, a 'combative' rather than 'collaborative' styles of learning, an emphasis on high pressure, timed formal examinations for assessing students and differences between colleges in the mixture of male and female students admitted.

Different levels of performance by male and female students could be the result of differences in their approach to studying: the Approaches to Studying Inventory (ASI) (Entwhistle and Ramsden, 1983) is widely used in research in higher education to measure undergraduates' approaches to studying. Although this measure has been applied in a number of research studies, Richardson (1993), reporting that most studies of approaches to studying had not considered gender differences, used the ASI to measure the approaches to studying of two samples of undergraduates studying social science at Brunel University and found that neither sample provided evidence of a difference in the approaches used by male and female students. A later study (Hayes and Richardson, 1995) compared the scores for two sub-scales of the ASI between male and female students studying for arts and science degrees at three Oxbridge colleges with different mixtures of male and female students. One college's students were all female, another had equal numbers of men and women and in the third, men outnumbered women by 2:1. In this study there were differences between men and women in their scores for two sub-scales of the ASI which varied according to subject and college. The distinctive samples studied in these two research papers (Richardson, 1993; Hayes and Richardson, 1995) and their respective findings: of no sex differences in one context and differences of varying degrees in another, make it difficult to form general conclusions. The samples described in these papers were highly selected and were drawn from distinctly different institutions; in the first case by subject, and social science students may exhibit fewer differences between male and female approaches than students pursuing other disciplines. In the second case, students in a range of subjects are included, but the broad grouping of arts and science subjects means that within these groups, the distribution by subject may vary between male and female students or from college to college and so differences in ASI scores cannot be attributed to either gender or college with any certainty. By themselves, these two studies make only a small impression on the gap identified by

Richardson (1993) and raise the question of how differentials might vary from one institution to another. A further setback is that in a later study, Meyer (1995) argues that the development of the ASI, which was not designed to distinguish between male and female students' responses, resulted in a measure which is unable to detect variations between the sexes in their approaches to study.

Considering assessment in general, Gipps and Murphy (1994) highlight the need to consider the effects on performance of students' attitudes and approaches and the contributions, direct and indirect, of teachers and assessors to the outcomes of assessment. Academics, individually and collectively can influence the outcomes of assessment in several ways: combining the roles of teacher and examiner, they define the nature of the subject and 'success' within it, choose the method of assessment, determine its style and content and mark the work which students submit. Differences in the assessed performance of sub groups of students can arise from differences in the responses of assessors to the students and their work or from differences in the responses of male and female students to their lecturers or the assessments they set. In studies of performance at secondary level, the method of assessment has been shown to influence the difference in performance between boys and girls (Gipps and Murphy, 1994), with girls benefiting from the introduction of coursework assessment at GCSE level. The literature on assessment in higher education has tended to focus on encouraging lecturers to adopt innovative forms of assessment rather than researching the impact of changes in assessment on the performance of sub-groups of students (see for example, Brown and Knight, 1994) Kniveton (1996) found that women students responded more positively than male students to statements that continuous assessment was fair, reliable, able to measure a range of ability and allowed students to organise their own work pattern. As yet, there appears to be little published research comparing the performance of undergraduates on different types of assessment. Upward trends in degree class have been attributed to the increasing use of coursework (Gibbs and Lucas, 1997). If this is the case, then we might expect women to have benefited more from the upward trend in degree class than men.

Outcomes of assessment may also differ between men and women students if assessors respond differently to work produced by male and female students, or if decisions taken by examiners are more likely to be favourable if the student is one sex rather than another. Repeated attempts to investigate these possibilities have lead to different conclusions, even after repeated analysis of the data (Bradley, 1984 and 1993; Newstead and Dennis 1990; Dennis and Newstead, 1994).

2.3.2 Age on entry to higher education

Steady increases in the numbers of mature students admitted to universities in the UK have resulted in a large minority of students being admitted to first degree courses at the age of 21 or above: so that by 1993/94, 15% of full-time, home undergraduates in the UK were aged 21 -24 and a further 20% were aged 25 or over on entry to higher education (CVCP, 1995). As the participation of mature students in higher education increased, researchers began to investigate the differences between mature and younger students in their experience of being a student, attitudes to studying, and academic achievement and completion rates (Metcalf, 1993). Initially there were concerns that mature students, returning to education after some time, might lack study skills or be at a disadvantage in subjects such as science and technology, where recent innovations might have occurred (Richardson, 1994). This proved not to be the case: as several studies carried out in the nineteen seventies and eighties (see for example, Bourner &Hamed, 1987; Woodley, 1984; Walker, 1975) concluded that mature students' degree results compared favourably with those of younger students. Some of these comparisons between mature and younger students (Peers, 1994; Walker, 1975; Brennan, 1986) were limited by the small numbers of mature students contributing data, particularly in studies of performance in 'old' universities, where fewer mature students were admitted. Similarly, a review of research carried out by Richardson (1994) was largely based on studies using data from the period when mature students were still a small minority.

The small numbers of mature students contributing data may explain why most researchers were reluctant to adopt a multi-factorial approach in their analyses. Peers (1994) found age to be a statistically significant factor, other things being equal, in determining degree class in science and technology, but not in other subjects, but acknowledges the potential lack of power associated with very small numbers of mature students. Research carried out using more recent data, and based in 'new' universities, in which mature students are represented in larger numbers, shows that performance on degree courses initially improves with student's age on entry to the course, until, for students entering degree courses in their late thirties or early forties mean performance falls, though not to below the level achieved by traditional students (Bourner and Hamed, 1987; Simonite, 1997).

Mature students differ from traditional students in several ways: they are more likely to be admitted on the basis of vocational and other 'non-standard' qualifications and those with A-levels or Scottish Highers tend to have lower grades for these qualifications than traditional students (Metcalf, 1993). Mature students also include a higher percentage of women and part-time students than students in the traditional age group and are distributed differently by subject. Changes in the distribution of these characteristics amongst mature students could lead to changes in the differences in the performance of students in different age groups, but, in the absence of multi-factor analyses, these changes cannot be anticipated. The lack of multi-factor analyses also means that the effects of changes in teaching and assessment in higher education also cannot be anticipated since, for example, the extent to which the differential between mature and traditional students depends on methods of assessment or the rewarding of transferable skills is unknown.

Studies of mature and other students' approaches to studying have shown (Richardson, 1993) that mature students are more likely to adopt a deep approach to studying than younger students. Sub-scales of the Approaches to Studying Inventory (Entwhistle and Ramsden, 1983) measuring behaviour associated with a deep approach have been shown to be positively correlated with age, while those associated with a surface approach to learning are negatively correlated with age (Richardson, 1995). Although these findings give support to the theory that the mature students' success is explained by their deeper approach to learning, no relationship was found between ASI sub-scale scores and degree class. This could have been because of the time elapsed between the measurement of ASI and the assessments on which degree classifications were based. However, another analysis failed to find a relationship between ASI scores and A-level qualifications so the problem may be that ASI does not measure those approaches to learning which are most related to success.

2.3.3 Entry Qualifications

Universities use A-level grades to select students, but the research literature shows an additional expectation, that these grades will predict students' subsequent academic achievement (for example, Sear, 1983; Johnes, 1992; Bourner and Hamed, 1987; Peers, 1994). The difficulties associated with using degree class as a measure of academic achievement were discussed earlier, in section 2.2, and similar problems occur in measuring performance at A level. In most studies, A-level scores are calculated by assigning points to each grade (E=2, D=4, C=6, B=8, A=10) and calculating each candidate's total score. The calculation of A- level points assumes comparability between boards and subjects and different A-level syllabuses, allocates an arbitrary set of values to each grade and adds a student's scores for each A-level taken without weighting, so that all subjects contribute equally to the final score, however relevant or otherwise they may be to the student's degree course. The 'points' system also means that a student with high grades for two A-levels may achieve the same number of 'points' as a student with low grades for three A-levels, which may not be appropriate. Some authors solve this by considering students with 2 or 3 A-levels separately (for example, Bourner and Hamed (1987)).

Studies designed to measure the association between A level grades and degree outcomes have consistently found only a weak relationship between A-level grades and degree results. Peers and Johnson's (1994) meta-analysis of the findings of 20 published studies produced a combined estimate of the correlation between degree results and A level grades of 0.276, suggesting that A-level scores account for approximately 7.6% of variation in degree outcomes. In some ways, the low correlation between A levels and degree outcomes is unsurprising: students with the same achievements at A-level may respond differently to the learning environments offered to them as undergraduates and assessments at degree level are unlikely to measure the same abilities as are measured at A- level, even where similar subjects are studied.

Some studies attempting to measure the relationship between degree results and entry qualifications will have been affected by a lack of comparability between the degree results obtained in different subjects or institutions. This problem could be tackled by using a multilevel model including institutional and/or subject variation, but as yet this does not appear to have been attempted.

In studies that control for the effects of other factors, some variations, according to either the context or students characteristics, have been found in the relationship between degree outcomes and A-level grades. Hence Sear (1983) found that the relationship between degree class and A-level scores appeared weaker amongst mature students than for students as a whole; Bourner and Hamed (1987) found, on CNAA degree courses, that the relationship between degree class and A-level grades appeared to vary between subjects, being strongest for languages and arts and weakest for engineering and technology and health-related subjects. Since other potentially influential factors were not controlled, interpreting this result is problematic.

More complex analyses, by Peers (1994) and Johnes (1992), examined the impact of A levels on degree outcomes and confirmed that A-level scores explain some variation in outcomes, but that substantial variation remains to be explained by other factors. Peers' (1994) analysis of the degree performances of students at the University of Manchester was designed to show variations in the relationship between degree outcomes and A-level scores according to age and gender. Carrying out separate analyses in each of four subject areas (humanities, social science, engineering and technology and physical sciences), Peers found that A-level scores contributed significantly to degree class after allowing for age and gender, but did not present parameter estimates which could be used to estimate the size of this effect. Peers' (1994) discussion of these findings emphasises the role of A-level grades in the admission process: his concern is to determine whether, after allowing for A level points, degree outcomes depend on factors such as age and gender, holding the view that if this was so, then an admission process based solely on A-levels would be unethical, and would discriminate against students belonging to those groups which achieve better outcomes than others with the same entry qualifications. Johnes (1992) uses her analysis to argue that widening access to higher education by lowering admission criteria would lead to a higher proportion of degrees being awarded in the lower degree classes. These arguments are speculative, as both Johnes and Peers are extrapolating their findings to predict the outcome of admitting students with entry qualifications outside the range covered in their sample: Johnes, by imagining what would happen if entry criteria were to be lowered and Peers by suggesting that entry criteria should be lowered for students of the appropriate age and gender.

Although alternative routes into higher education are recognised, A-level grades are likely to continue to be used to select students. Wood (1991) describes an unsuccessful attempt, made in the 1970s, to improve on the lack of predictive power of A-levels by using an aptitude test. Wood (1991) reports the relationship between achievement in higher education and aptitude scores as similar to the relationship between achievement in higher education and A-levels. After controlling for A-levels, aptitude test scores did not contribute significantly in analyses of subsequent academic achievement. Wood interpreted this finding as an example of the general case that aptitude tests are poor predictors of those abilities which they are intended to predict.

One element in the expansion of the higher education system has been the admission of students with qualifications other than A-levels. Bourner and Hamed's (1987) study of CNAA graduates found that students with non-traditional qualifications obtained marginally higher results than those with A-levels. Leopold and Osborne (1996) found that former Access students were almost as successful as students with 'school qualifications' although it should be noted that no steps were taken to control for the differences between groups in their distribution by subject studied. Comparisons between students with traditional and other entry qualifications are complicated by differences in the age distributions of these groups, with students with alternative entry qualifications tending to be older on average than those with A-levels (Bourner and Hamed, 1987). Comparative studies of students with traditional and alternative entry qualifications are few and those that exist are relatively small. Contributing factors include the small numbers of students for whom degree outcomes are as yet known and of defining groups which are sufficiently homogeneous to justify further study and the difficulty of classifying students' sometimes varied educational history.

In analyses of outcomes at degree level, entry qualifications are of interest not only because of their role in the admission, but also as a way of adjusting for pre-existing differences between groups whose performance in higher education is to be compared. In this context, the weak relationship between entry qualifications and degree outcomes means that other controlling variables will be needed to explain a high proportion of variation in the academic achievement in first degree courses.

2.3.4 Class size

Mahler et al (1986) observed a large number of classes of different sizes and found differences in the cognitive diversity of teaching and the characteristics of verbal interactions occurring in small, medium and large classes. In this study, classes with 17-50 students were designated large; hence the maximum class size studied was substantially lower than those which occur commonly in higher education. The impact of class size on academic achievement is an important issue, because assessment is intended to measure student achievement, not institutional resources. Class size might influence student's learning and achievement directly, as a result of the student being part of a large class, or indirectly, by producing changes in the teaching style or number and type of assessments used. In either case, this means that the assessment is measuring the impact of class size as well as student performance. Where students aiming for the same award can choose between programmes, this would mean that some students might gain an unfair advantage by virtue of taking a programme including more classes with small enrolments (if large class sizes are associated with poorer outcomes).

Published evidence on the impact of class size on academic achievement in first degree courses appears to be based entirely on the findings of a series of papers based on termly module reports produced for the Modular Degree Programme at Oxford Brookes University (Fearnley, 1995; Lindsay & Paton-Salzburg, 1987; Gibbs & Lucas, 1995; Gibbs, Lucas & Simonite, 1996). In these papers, the impact of class size is measured by studying the relationship between the number of students enrolled on a module and the module average (Fearnley, 1995; Gibbs & Lucas, 1995) or grade distribution (Lindsay & Paton-Salzburg, 1987). The number of students enrolled on a module has some limitations as a measure of class size, since teaching can be arranged in different ways. For example, students taking a module with large enrolment may attend lectures as one group but be divided into sets for practical or seminar work. In very large modules, say with enrolment over 200, lectures may be repeated if student numbers exceed the capacity of the largest lecture theatre available. Where students have the opportunity for individual contact with staff, then access to staff will depend on the relationship between time available and student demand rather than simple class size *per se*. Modules with similar enrolment may therefore teach students in groups of different sizes and offer students different opportunities for individual attention.

Fearnley (1995), Lindsay & Paton-Salzburg (1987) and Gibbs and Lucas (1995) all look at the relationship between the module level results and class size, after controlling to some extent for subject differences in mark distributions by running separate analyses for subsets of modules within different subjects (Fearnley, 1995) or within broad subject groups (Lindsay & Paton-Salzburg, 1987; Gibbs and Lucas, 1995). All three studies use data for several years, but treat repeated observations on the same modules as independent. This approach was criticised by Bristow (1989) since a more appropriate, longitudinal approach would have been able to relate changes in module enrolment to changes in the outcomes, providing a more direct evaluation of the effects of class size. Bristow (1989) also put forward the view that differences between modules needed to be adjusted to take into account any pre-existing differences in the abilities of the students. Modules that have a number of pre-requisites, are specialised in nature and are not compulsory are likely to have lower recruitment than modules at the same level which are more easily accessed or which are compulsory for some students. These modules are also more likely to be taken later in a degree programme. Higher mean marks for smaller modules could therefore be explained as the result of students enrolled on specialist modules being more motivated, more experienced or more able in the relevant domain.

The studies of class size provide evidence of a negative relationship between module outcomes and class size but the measurements of this effect are unreliable because of the narrow focus of analyses which exclude other performance related variables. In addition the use of module level data offers no opportunity to examine the impact of class size at individual level nor of how this may vary from one student to another.

2.3.5 Method of assessment

Changes in higher education in the United Kingdom have lead to the introduction of new forms of assessment, particularly through the growth in the continuous assessment (Jackson, 1996). Diversity in assessment has been criticised as threatening the reliability of student's degree classifications (Morrison et al, 1997) and used to support the case for the abolition of degree classifications (MacFarlane, 1998). The impact of new forms of assessment on students have been studied largely through attitude questionnaires (for example, Kniveton, 1996; Franklyn-Stokes and Newstead, 1995) rather than studies of the effects of assessment methods on students' marks or grades or the role of assessment methods in determining the differences in achievement between sub-groups of students. Gibbs and Lucas' (1995) study of module averages at Oxford Brookes University showed that in modules with a high percentage of coursework assessment, module averages were higher than in modules relying on traditional examinations. Gibbs and Lucas' (1995) work has been used to explain the upward drift in degree classifications which occurred in the 80s and 90s (see for example, Elton, 1998, Chapman, 1996). Because Gibbs and Lucas' (1995) study used data recorded at module level, there were no opportunities to see how different sub-groups of students were affected by the method of assessment (such studies being more common at secondary level). Although the widespread use of coursework and its impact on degree classifications is widely recognised, this recognition has not lead to more detailed work on its effects on issues such as gender or subject differences. This is surprising, because the relationship between assessment method and outcomes is important when students in a modular scheme are able to exercise a degree of choice in constructing their programme of studies. If students studying for the same award may follow programmes of modules in which the proportions of coursework assessment differ, neither student should be disadvantaged by choosing one acceptable programme of modules than another. Assessors in modular courses must therefore seek to eliminate systematic differences in the outcomes of assessment by different methods.

2.3.6 Subject studied

Annual statistics provided by HESA (for example HESA, 1998) show that subjects differ in the proportions of degrees awarded in each class. Tomlinson and MacFarlane (1995) found gender differences in degree classifications were partly explained by male and female graduates having studied different subjects. In the past, differences between subjects in the class of degrees awarded have been accepted within higher education: the external examiner system is designed to tolerate subject differences, since external examiners are asked only to ensure that institutions apply comparable standards *within subjects* (CVCP, 1984).

Jackson (1996) describes modular courses as highlighting the differences in marking practices between different subjects. This is particularly the case within modular schemes covering a wide range of subjects, and when students combine degree subjects. Variations in mark distributions are evident at module level: Yorke et al (1996) studied the means and standard deviations of marks generated in 8 subjects within 6 institutions. Their study found evidence of consistency between institutions in the rank ordering of subjects according to the means and standard deviations of marks awarded. Bridges et al (1999) performed a similar study, leading to the claim that "marking behaviour is a potential cause of inequity of outcomes for students who are given choice over the combination of elements, modules or subjects". In both cases the authors attribute subject differences in mark distributions to the attitudes of assessors within the relevant disciplines. These studies demonstrate the existence of subject differences in mark distributions at module level, but, in both cases, large numbers of marks have been processed, without regard to the marks having been awarded within a limited number of modules or to individual students having contributed multiple marks. The effects of other performance-related variables are not considered.

Section 2.2.6.5 explained the need for comparability between subjects to be accepted in the context of analyses of academic achievement covering several subjects. As measures of achievement at module and degree level depend on the subject studied, such analyses also need to take into account the differences between subjects in the distribution of marks or grades. In the models applied in later chapters, fixed effects and complex variance structures will be used to represent the between subjects differences in mark distributions.

The next section discusses the modularization of higher education in the United Kingdom and its implications for the assessment of undergraduates in more detail.

2.4 Modularisation in higher education in the United Kingdom

The expansion of higher education in the UK in the nineteen-eighties and nineteen-nineties and the associated reductions in per capita funding lead to concerns about standards which an enquiry called the Graduate Standards Programme (HEQC, 1996b) was designed to address. Within this programme, a series of reports investigated the provision of degree programmes with a modular structure (see for example, HEQC 1996b). An important finding was that less than 10% of institutions providing first degree courses employed a linear curriculum framework, while 65% provided degree courses within a modular framework. The remainder provided degrees within a unitised curriculum framework (HEQC, 1996b). The use of unitised and modular frameworks was found to be a recent development (HEQC, 1996b); over 60% of higher education institutions had introduced a modular or unitised framework within the previous four years. Although modular courses share some common principles and structures, each institution produces it own detailed definitions, regulations for constructing programmes, rules governing progression and systems for aggregating the outcomes for accumulating individual modules.

A student who enters a modular degree course plans a programme of study that identifies the modules to be taken in each term or semester of their period of study. For a given award, the student will need to pass specified numbers of modules at each of two or more levels (within the CAT framework, 120 CAT credits are required). The number of modules which are required at each level, and their size and shape are determined by the institution. The content of the modules selected must 'add up to' a valid degree by conforming to the requirements of the degree award for which the student is studying. By controlling which modules can be taken, a university is able to control the breadth and depth of the course leading to the award.

Assessment takes place within modules and the outcomes are used to decide whether students can continue from one academic year to another. If a student passes a selection of modules which meets the criteria for the awarding of an honours degree, their degree classification is based on an aggregate measure of their performance in some or all of the modules taken. The system for classifying degrees is another feature of a modular degree course that is chosen by the institution.

Assessment within a modular framework raises some important questions: the aggregation of module results and the fact that alternative programmes of study can lead to the same degree award both involve implicit assumptions of comparability between modules at the same level. The larger the number of modules and the greater the degree of choice, the more extensive these assumptions are.

Within a modular scheme, programmes leading to degrees in different disciplines may have modules in common: modular courses have encouraged an inter-disciplinary programmes and the development of modules which serve several degree awards (Somerville, 1996). The economies of scale provided by this kind of teaching are one of the reasons cited for the rapid increase in modular courses (HEQC, 1996). Shared teaching may occur because the curricula for closely related subjects can overlap; for example, students of geology and environmental science may both need to study hydrology. Other sharing of modules can occur between students of apparently unrelated subjects, if students need to have reached a certain level of education in another subject in order to pursue their own studies - for example 'service' modules in mathematics may be taken by students taking any of a wide range of business, science, engineering and technology degrees. Modules that are shared by a number of degree awards lead to assumptions of comparability between modules in different subjects. Questions of comparability within modular courses are discussed in detail in section 2.4.1.

Related to the question of comparability are the roles of internal and external examiners, which have been subject to substantial changes as the result of modularisation (Silver et al, 1995). As a result of these changes, for the majority of students graduating within a modular framework, degree classification is determined 'by formula' according to the system chosen by the institution (Silver et al, 1995). Aggregation methods differ in the way that they reward different patterns of achievement (Wood, 1991) and in the reliability of the aggregated score, so the choice of aggregation system is an important one.

In modular degree courses, as a student accumulates the academic credits required for the award of a degree, they are provided with information about their results after each round of assessments. This information allows the student to judge how well they are performing and informs their choice of the rest of their programme. As the student draws towards the end of their programme, they are able to calculate the performance required in their remaining modules to achieve a certain degree classification. The feedback provides an ongoing message to the student, potentially influencing their motivation and performance, but the message conveyed by a given set of marks and grades depends on how, at the end of the programme, the results will be aggregated. The characteristics of different aggregation systems and their influence on students' motivation are discussed in section 2.4.2.

2.4.1 Comparability within a modular degree scheme

Within a modular degree course the aggregation of assessment data and the availability of a choice of programmes leading to the same award involve implicit assumptions of comparability. As for comparability between degree awards, statistical and subjective approaches to defining comparability between modules are considered and Cresswell's (1996) judgmental definition of comparability is used to discuss whether comparability between modules is achieved in practice. When a degree course allows students to choose between alternative modules, all permitted programmes of study are treated as having equal educational value, so there are implicit assumptions of comparability between the modules offered as alternatives. The aggregation of results also involves assumptions of comparability between modules. Several features of modular courses: choice of modules, different modes of study, intercalation, shared teaching, inter-disciplinary programmes and the aggregation of results, mean that comparability is assumed, *in practice*, between many groups of modules. Assumptions of comparability are made between groups of modules, but for simplicity, comparability will be considered between pairs of modules. The extent to which comparability between modules can be achieved within a modular framework and the implications for comparability between degree awards are discussed.

When students are able to make choices about the order in which to take modules in their programme, there is an underlying assumption that comparable standards will be used on each occasion. Variation in standards would mean that students taking the module on one occasion were disadvantaged compared to other students graduating in the same year who took the same module on different occasions. The assumption that, at the level of degree awards, standards are maintained over time also implies that standards are maintained between occasions within modules or other components of degree courses from year to year.

The assumptions of comparability between different modules are more complicated: many of the assumptions will refer to modules within the same subject or discipline, but to avoid the duplication of material in modules which will be taken by the same students, they will have been designed to cover different domains. Since different domains within a subject are valued and experienced differently by individuals, including internal and external examiners, it may be difficult to define equivalent standards in each domain (Goldstein and Cresswell, 1996).

Comparability of standards has not been required between degrees in different subjects, on the grounds that meaningful comparisons cannot be made (Silver et al, 1995). However, within a modular degree course, comparability may be assumed between modules in different subjects, because of 'service' teaching, a broad approach to defining the content of a degree, or because of combined honours degrees. There are two difficulties here: first of defining comparable standards for modules in different subjects and second the involvement of several examination boards, each dealing with modules in a different subject, in deciding or confirming a student's results.

The co-existence of modules and degree courses in different subjects within large modular schemes, has focused attention on the traditional differences between subjects in mark distributions (HEQC, 1996), with some support for a 'norming' approach to comparability shown in calls for the distributions of module marks in different subjects to be more similar. Yorke et al (1996) found that differences between subjects in marking practices at module level were stable across a number of institutions providing modular degrees. It is interesting to see that modularisation, has simultaneously created support for stricter norm referencing and a trend for assessment to become more criterion referenced (HEQC, 1996). There seems to be some feeling that reducing the variation between subjects is a desirable goal (Elton, 1998; HEQC, 1996). In practice changes would be unlikely to be perceived as 'fair' - even supposing a 'standard' distribution of marks/grades were agreed, there would be discontinuities between results before and after the changes and differential effects on students in different subjects, creating different effects for male and female, or mature and traditional students which would be perceived as unfair. This poses a dilemma for managers of modular degree schemes where the current position may also be perceived as unfair, if students following a programme leading to a degree in one discipline are more likely to be awarded a first class or upper second class honours degree than other students who are studying within the same modular scheme for a degree in another discipline. Requiring comparability of standards between modules in different subjects leads to the demand, previously avoided, for comparable standards to be achieved in degrees in different subjects.

2.4.1.1 Statistical approaches to defining comparability between modules

The 'norming' definition of comparability between modules, which requires the distribution of marks or grades to be the same for both modules, would introduce an extraordinary degree of uniformity over the whole of a modular scheme, by requiring the same distribution of marks or grades in all modules. Even within disciplines, there is no good reason why modules should produce the same grade or mark distributions, since there will be differences between modules in intake as a result of students choosing different programmes.

One aspect in which comparability between modules differs from comparability between degree awards is that where two modules must be comparable in standards, there may be a group of students who have taken both modules. The subject-pairs method, is designed for assessing comparability in this situation, but, as with other statistical techniques, work on public examinations has identified a number of problems (Goldstein & Cresswell, 1996). Detailed criticisms of the subject-pairs technique are given by Newton (1997) and Goldstein & Cresswell (1996). One problem is that, in many cases students who have taken both modules may not be representative of the intake for either of the modules. For example, if one of the modules provides an inter-disciplinary contribution to a range of particular degree award while the other is a 'specialist' module recruiting students from a narrower range of awards. A second problem is that some students will receive higher marks or grades for one module while others will receive higher marks or grades for the other. Any adjustment of marks or grades intended to achieve comparability over all students will inevitably make the outcomes less comparable for some. Thirdly, the subject-pairs method is based on the assumption that assessment in both modules measures the same underlying attribute (uni-dimensionality). As modules are specifically designed to have non-overlapping syllabuses, this is uncertain. The next section considers how comparability between modules can be defined in terms of academic judgement rather than by statistical means.

2.4.1.2 A subjective definition of comparability between modules

A potential definition of comparability between modules, based on Cresswell (1996) is that:

> *'two modules have comparable standards if candidates for one receive the same marks or grades as candidates for the other whose assessed attainments are accorded equal value by examiners and other assessors accepted as competent to make such judgements by interested parties within and outside the institution'*

Defining comparable standards of achievement between different modules will be challenging since even within subjects, module syllabuses are designed not to 'overlap'. The definition of comparability above does not attempt to equate attainment in different domains - only the *values placed on those attainments* by the awarders (Cresswell, 1996). In theory, this allows meaningful statements about the comparability of modules to be made even when they are in different subjects, but in practice it may be difficult for the awarders to attach values to attainments in subjects outside their own area of expertise or to be accepted as competent to do so by interested parties.

'Examiners and other assessors' are defined as the relevant internal and external examiners, plus other members of the committees or examination boards which have responsibility for awarding and confirming module marks. In most cases, the 'interested parties' who need to accept the awarders competence will consist of students, staff and administrators within the modular scheme, but may also include past and future students. The accreditation of certain courses by professional associations and the awarding of CAT points for successful completion of modules identify interested users outside the institution.

Having defined comparability between modules, the next step is to consider how this can be achieved. Initially, standards are partially defined either when a new module either is designed specifically for a degree award, or when an existing module is considered for inclusion in programmes leading to a degree award. At this stage, there will be opportunities to evaluate the educational content of the module and the nature of the assessment and the intended learning outcomes, relative to other modules that may be taken.

Later, on each occasion that a module is taught, standards are further determined by the teaching staff's interpretation of the syllabus, the writing of assessments and marking schemes, the marking of students' work and when the reviewing and confirmation of results by the examiners. A common way for universities offering modular degree courses to organise their exam boards is in a two tier system in which a subject examination committee discusses the outcomes of individual modules, while a final award committee discusses borderline cases or cases in which students circumstances need to be taken into account (Silver et al, 1995; HEQC, 1996). Subject external examiners are involved in the first tier, discussing module outcomes within their subject and reviewing the progress of students enrolled on degrees wholly or partly based in the external examiner's subject. The second tier involves one or more chief external examiners and considers the final awards of students across the whole modular scheme. The purpose of this committee is to see that the assessment process deals fairly with individual students whose circumstances need to be taken into account and to define the final awards for borderline cases. Only one group of examiners and other assessors is involved in confirming the marks or grades awarded in modules within the same subject or on different occasions. These examiners will be familiar with the syllabuses of each module, have been responsible for setting the assessment and students' assessed work will be available to them. Since all concerned will be expert in the areas of the curriculum covered by the modules, their competence is likely to be accepted by all the interested parties. The problem they face is in valuing attainments in different domains, which may have been assessed in different ways.

For comparisons of standards between modules in different subjects, interested parties within the institution will comprise staff, students and other relevant groups, but there will be two groups of examiners and other assessors, one from each subject. Cresswell (1996) states that for comparability to be achieved, both sets of awarders must judge standards to be comparable and for each group of awarders this involves detailed work with reference to a range of materials from both modules. Within a two tier system of exam boards, examiners in one subject may not be able to contribute to the design of assessment instruments, or have access to scripts and other assessed work for modules which are the responsibility of another examination board. They may be unable to place a 'value' on attainments related to a subject outside their area of expertise. Where modules are assumed to be comparable because a student is taking a combined degree, there may be no information exchanged between the two (main) sets of examiners. If comparability between modules in different subjects is achieved within this system, then it is likely to be because interested users agree to it, in order that the modular scheme can operate, despite the lack of evidence that awarders are able to make or have made reliable judgement.

Within a modular scheme, comparability is sometimes achieved through painstaking and detailed consideration of the design of modules and assessments and of students' assessed work and sometimes by a mutual agreement by those concerned to accept module credits, marks or grades 'at face value'. There are two ways in which this concerns research studying student achievement within a modular degree course: first, the findings of the research must be interpreted in a context in which the same definition of comparability is accepted and second, the research may be used to influence the achievement of comparability between modules, by providing empirical evidence which can be used to maintain the agreement to accept modules as comparable. Analyses of module outcomes within modular schemes, particularly those that compare performance between modules after controlling for differences in their student intake, will be generating valuable information for ensuring that the system of assessment within modules is being operated fairly.

2.4.2 Systems for classifying honours degrees in modular degree courses

The aggregation of students' marks or grades for different modules is one of the reasons for the widespread assumptions of comparability made within modular degree courses. This section shows that aggregation systems have important implications for students as they progress through their programme of studies and on graduation. Systems for combining module results to make decisions about progression or graduation vary between institutions. Degree class is determined 'by formula' for the majority of students who follow a modular degree course, according to the system adopted by the institution and it is important that the characteristics of alternative systems are clearly understood. The proportion of modules taken which contribute to degree class, the method for combining module marks or grades and the boundaries used to classify the combined mark or grade are important aspects of systems for determining degree classifications. Related to these are the definitions of borderline candidates whose results are referred to the second tier of examiners and the procedures for determining which class will be awarded.

2.4.2.1 Modules contributing to degree classification

The HEQC (1996b) reports variation between institutions in the selection of modules on which degree classifications are based, with some institutions aggregating the results for all modules passed at the appropriate level and others aggregating the results for only a proportion. When a given number of module passes are required for an honours degree, the most obvious effect of excluding a student's lowest marks is to produce a more favourable aggregate mark and in some cases a more favourable classification. A second effect is a reduction in the reliability of the overall result. Since assessment is not an exact science, module marks and grades are subject to measurement error. Any reduction in the number of modules contributing to degree class reduces the reliability of the result, since the reliability of an aggregate measure is inversely related to the number of items combined. The use of a students' 'best x' results to decide their degree classification, where x is the number of module outcomes combined, increases the probability that the combined mark will over-estimate the student's ability since module marks with positive errors are more likely to be included. This effect will be particularly marked if a student has taken more than the minimum number required at this level, since their 'best x' results represent a smaller proportion of their achievements. The use of some rather than all module marks to determine degree classifications reflects a different view of what degree class is intended to indicate. Warren Piper (1986), quoted in HEQC (1997), detected two philosophies amongst aggregate schemes: one intending that degree class should summarise all a student's assessed work and the other intending it to reflect a student's best assessed work. The extent to which these two approaches will place a student in different classes depends on the variability in the student's assessed performance: the greater the variation in the student's marks, the more likely they are to be classified differently. The decision to base degree classifications on the student's best assessed work (at the appropriate level) is one which rewards students whose performance is erratic at the expense of steadier performers.

2.4.2.2 Methods for combining module outcomes

The HEQC's (1996b) survey of institutions providing modular first degree courses identified just one institution operating a grade points system with the remainder aggregating module marks. Good and Cresswell (1988) concluded that combining marks is preferable to combining grades, although Cresswell (1988) showed that the loss of information associated with using grades can be minimised by increasing the number of grades used to record achievement on each module. An advantage of combining grades is that consistent classifications are produced for students who have achieved different marks but identical grade profiles; this is important if student transcripts include module grades but not marks. At present, no detailed information about the content of student transcripts provided within modular degree courses appears to be available, however, the Dearing Report recommends universities adopt a common format for student transcripts, which is expected to include individual marks (NCIHE, 1997).

Amongst the universities and other institutions surveyed by the HEQC (1996b) which were not using a grade points system, a variety of approaches to weighting the marks for modules at different levels were in use, but the aggregation systems were not explicitly stated. Morrison et al (1997) suggested that a wide variety of aggregation systems may be in operation, although Wood (1991) 's description of the practice of combining marks by addition as "so entrenched that it would come as a great surprise to many to be told that it does not necessarily have to be done this way" suggests that there could be widespread use of averaging (weighted or otherwise) to decide degree class.

Averaging allows a good mark in one module to make up for a poor mark in another, enabling students to make up for occasional poor performance and increasing the reliability of the aggregated mark by allowing positive and negative errors of measurement to cancel each other out. One concern which has been expressed regarding this compensation is that it may exacerbate the differences in marking practices between disciplines (HEQC, 1996b). The tendency for scientific and technical subjects to generate more extreme mark distributions is well known. If an aggregation system allows compensation, it is possible that students in subjects with extreme mark distributions are more likely to have their degree class improved as the result of one or two very good results than students in subjects with less extreme mark distributions. Equally, students in technical and scientific subjects may be at greater risk of having a lower degree class, as the result of a poor result, although the use of the 'best *x*' results to decide degree classifications would protect them. In the absence of published evidence of the variability within students mark profiles, these arguments are speculative. For students with the same mean but different levels of variability, averaging the marks, produces the same degree classification for 'steady' and 'erratic' performers if all the relevant marks are combined but favours the erratic student if only the 'best *x*' marks are considered.

Marks can be combined in other ways; Morrison et al (1997) applied a grade points system, an 'arrangement system' and the universal marks system to a hypothetical mark profile covering six modules and compared the results. The hypothetical profile consisted of marks graded A, A, A, C, C and Fail and a different degree classification was obtained from each system. The sets of marks used were constructed to be particularly challenging: this is a useful approach for detecting anomalies produced by systems of aggregation but provides no information about the results these aggregation systems might produce when applied to actual student records. Although it is important to recognise that different aggregation systems may lead to different outcomes, these differences represent less of a problem than Morrison et al (1997) suggest. The variation in outcomes is produced because the three systems respond differently to the pattern of achievement presented. Different responses are most likely when a student's mark profile is extremely varied as is the case in the example provided by Morrison et al. (1997). If the student had performed consistently, the three aggregation methods would be more likely to agree. A key element in choosing a system for aggregating marks was described by Wood (1991) as "a principled consideration of what kind of achievement is to be rewarded". It is important that a system of aggregation accurately reflects the intentions of the university in the way that it values different patterns of achievement. It must also be transparent, so that students are aware of the system which will be used to determine their own degree classification.

Grade point systems, one of the methods tested by Morrison et al (1997), have not been taken up widely in modular degree courses as the HEQC (1996b) reports just one institution operating this kind of system. Grade points systems operate in a similar fashion to averaging, but converting from marks to grade points produces some loss of information.

'Arrangement systems', also tested by Morrison et al (1997), appear to be used mainly in borderline procedures. Arrangement systems are based on the idea that for a student to be awarded a certain class, they must achieve a minimum number of grades at that level (or higher). Awarding the highest grade achieved in a majority of modules is an example of an arrangement system. More complex arrangement systems may specify in more detail the distributions of grades which qualify for an overall grade, for example by fixing the number of grades which must be no more than a certain distance below the overall grade. With an arrangement system, an uncharacteristically poor grade may not change the overall result, and similarly, one or two grades above the student's usual level of performance will not raise their overall classification. In effect, 'arrangement systems' do not offer the degree of compensation which is available with averaging and one result of this is that widely varying grades may produce a low degree classification for a student's whose average lies within a higher class.

Morrison et al (1997) favour the universal marks system, partly because it facilitates transfers between institutions in the United Kingdom and the rest of Europe and partly because it performs well when combining marks achieved in modules for which the pass/fail and grade boundaries vary. The latter property is not particularly relevant since in practice variations in pass/fail and grading thresholds are widely recognised as undesirable within modular courses offering students a degree of choice.

2.4.2.3 Degree class boundaries and borderlines

There is general agreement between institutions of the boundaries for first, upper second and lower second class degrees (HEQC, 1996b). In most institutions, students are awarded first class, upper second class and lower second class degrees if their aggregated percentage mark falls in the intervals 70 and over; 60-69 and 50-59 respectively with greater variation between institutions in the thresholds defining third class and Pass degrees (HEQC, 1996b).

The HEQC's (1996b) report on academic standards in modular frameworks defines borderline candidates as students whose aggregate mark falls just below a degree class. In the absence of more detailed information, it seems likely that the aggregate mark referred to here is a weighted or unweighted average. Typically, the width of the borderlines varies from one class to another, with borderlines defined more generously at the first class threshold than at others (HEQC, 1996b). Although some institutions have no formal definitions of borderlines, the HEQC (1996b) reports a majority of institutions as having a standard mechanism for awarding a higher class than is indicated by the usual aggregation system. This finding is supported by descriptions by Silver et al (1995) and Adams (1996) of the treatment of borderline cases within the two-tier system of examination boards. As borderline degree classifications are also increasingly determined by formula, the characteristics of the procedures used at this stage also need to be clearly understood. Typically a borderline candidate is awarded a degree in a higher category if a large enough proportion of modules are in the higher class, with a variety of definitions of the modules to be considered at this stage: alternatively, some institutions use the mark achieved in a final year dissertation or project or in synoptic modules to decide whether to award a higher classification than is indicated by the aggregate mark (HEQC, 1996b). In effect a second aggregation system is applied to a student's marks and if this produces a higher class than the original aggregated result, the student is awarded the higher of the two classes.

Since it was established earlier that aggregation systems reward different patterns of achievement, the variability in a students mark profile will determine whether the second aggregation system applied to borderline candidates produces a more favourable result than the first. It has now been established that when universities choose key aspects of their system for determining degree classifications, the impact of these choices depends on the patterns of variation that occur within the profiles of marks and grades to be combined. As yet, there appear to be no published studies of the variation within students' performance on modular degree courses.

2.4.2.4 Interpretation of feedback during programme

When, students' records are updated after each assessment period, each student is able to review what they have achieved so far and think about what they can expect to achieve in the future. Their future choice of programme, their approaches to studying and hence their future record will be shaped by this process. Different students will respond to the same feedback in different ways; for example a poor result may spur one student into greater effort while another would accept the occasional poor mark and continue as before. Students who appear to be heading for a borderline may respond differently to their feedback than students whose record indicates they will finish within the middle of a degree class. A student's response to this feedback will also depend on the institutional system for deciding awards: if students know that only their 'best x' results will count towards their degree classification, they will be more relaxed than if all their module marks will be used. The method for combining the marks will also influence the relationship between a student's past and their future records at each stage in their degree, as the following example illustrates.

Example: Interpreting module marks achieved partway through a programme under different aggregation systems

This hypothetical example compares the implications of a particular set of marks when results are to be combined using different methods. Two systems for combining marks will be compared; the average mark and the best grade achieved in a majority of modules. The future marks required for the student to achieve a given degree class are compared between these two systems.

Suppose that 7 module marks will be combined to classify students and that the thresholds used to define first, upper second, lower second and third class are: 70%, 60%, 50% and 40% respectively. A student has achieved the following marks in four modules: 54%, 56%, 57%, and 62%. What must this student achieve in the next three modules, under different aggregation systems, for their final mark to fall in each class?

To achieve a first class degree:

If marks are to be averaged, the student needs to achieve an average of 87 in their remaining three modules. This is possible but unlikely given the student's record so far.

If the student is to be classified by 'majority grade', a first class degree is impossible at this stage since it requires 4 out of 7 grades to be in that class.

To achieve an upper second class degree:

With averaging, the student needs an average of 63.7 in the remaining modules; this is higher than their current average, but might be achieved in practice. With 'majority grade', the student must achieve three marks in the 2:1 category to be awarded a 2:1, so that for example three marks of 60 would be sufficient. This means that a 2:1 can be achieved with lower marks than those needed for an average in the sixties, as long as none of the marks falls to 59. For example, using majority grade, 61,62 and 63 will secure a 2:1 but 56, 67 and 68 will not.

To achieve a 2:2 degree

If marks are to be averaged, the marks for the remaining modules must average 43.7 over three modules. This appears relatively easy to achieve given the students past record, but in theory it is still possible for them to fail, for example if their remaining marks are 40, 40 and 50 (average 43.3), they will receive a third class degree.

With majority grade, as long as the student passes three further modules, they cannot fail to achieve at least a 2:2, since they have already achieved 4 out of 7 grades at or above the 2:2 threshold.

In this example the marks carried one message if degree class was based on average mark and another if degree class was based on majority grade. In each case, different patterns of future performance were required to achieve a degree in a specific class. The example shows the potential for a student's interim results to influence their future behaviour and performance, in a way that is mediated by the method of aggregation.

2.5 Summary and proposed analyses

This chapter has identified some areas in which the research literature on undergraduate achievement is limited and others where there is no relevant research evidence at all. Sections 2.2 and 2.3 showed that although studies in higher education have identified several performance-related factors, our understanding of the relative importance of these factors and of the interactions between them is limited by the small number of multi-factor analyses available. The use of degree classification to measure academic achievement has also been limiting and in some cases handled inappropriately.

The analyses that will be described in detail in chapters 4 to 6, based on the academic records of a cohort of students who graduated from the Modular Degree programme at Oxford Brookes University in July 1997, are designed to avoid the limitations identified earlier in this chapter. In these analyses, the effects of a number of student and module characteristics are fitted simultaneously, so that effects of each variable are estimated 'other things being equal'. Academic achievement is measured by the marks awarded to a student in each module taken within their degree programme. Module marks, measured on a scale from 0 to 100, are more discriminating than degree class and allow more detailed descriptions of the effects of performance related factors. The use of random effects models allows the variation between students to be measured within specific sub-groups, so that, for example, factors contributing to the greater dispersion in the degree of results for male students compared to female students may be identified.

Although the administration of modular degree programmes requires detailed, longitudinal records of students performance to be maintained, until now, the longitudinal nature of these data has not been exploited. The analyses presented later will therefore be breaking new ground. Longitudinal analyses allow the pattern of progress over time to be described and variations in the patterns of progress associated with particular subgroups of students to be explored. Patterns of progress may also vary between individual students and this too can be studied by random effects modelling. Section 2.4.2.3 showed how feedback after each assessment period could influence a student's future performance. We would therefore expect to find a variety of shapes amongst students' mark profiles. At any stage in the degree programme there will be some students whose feedback leads to a change in performance. The longitudinal analyses of students' records will provide the first information on these patterns.

The findings of later analyses will be discussed in relation to the features of modular degree courses that were discussed earlier in this chapter. Section 2.4.1 described how, in a modular degree course, comparability between modules is achieved by mutual agreement between examiners and interested parties rather than the application of statistical techniques. These agreements need to be maintained and supported by empirical data, consisting in its simplest form, of module averages and mark or grade distributions. Differences between modules in these statistics may indicate a lack of comparability, but could also reflect the effects of different student intakes. The analyses presented in later chapters will compare the results achieved in different types of modules, after adjusting for intake characteristics. These results will allow some of the assumptions within a modular framework to be evaluated.

The discussion of degree classification systems (in section 2.4.2) highlighted the contribution of the variation in a student's marks to the class of degree awarded. The analyses presented later will use complex variance structures to measure the variation within student records and to investigate factors affecting this variation. These findings will contribute new information on undergraduate performance, as previous studies provide no information on the variation within students' records.

Another way in which the analyses will add to existing knowledge is by showing the effects on individual students of factors whose effects on performance have previously been shown only in analyses of module level data. Sections 2.2 and 2.3 showed that while the introduction of coursework assessment is recognised as an important change in first degree courses, the extent to which coursework contributes to assessment in first degree courses has not been measured and there have been no major studies of the effects of different assessment methods on results measured at student level. This means that there is no evidence showing how changes in the methods of assessment have affected different groups of students, how the proportion of assessment based on coursework varies between degree courses or between programmes leading to the same award or of how this influences degree awards. Similarly, the effects of class size on students' achievements have been studied using data aggregated to module level. Longitudinal analyses of students' programmes will show how much the class sizes experienced vary within and between students' programmes and the effects of these variations on performance. The effects on performance of factors such as class size and assessment methods are of particular concern to the managers of modular schemes who need to provide evidence to support the assumption that comparable standards are applied in different modules. A multilevel, longitudinal approach will allow the effects of these factors to be studied more effectively than before.

Finally the longitudinal analyses may explain some of the findings of existing studies, by showing how the differentials in final degree awards identified in earlier studies are accumulated. For example, whether differences in the classes of degree awards are determined by differently shaped 'progress curves' or as the result of one group being more vulnerable than another to the effects of a heavy workload or responding differently to feedback of results during the course.

The use of a longitudinal, term by term approach to studying achievement within a modular scheme is new to the field of higher education. The analyses in this thesis will provide new information about the performance of students in first degree courses and improve our understanding of the findings of existing research. It is intended that the methodological approach used here will provide a model for the analysis of institutional data maintained by the modular programmes.

Chapter 3

The Sample

3.1 Introduction

This chapter describes how students within the Modular Degree Programme at Oxford Brookes University choose programmes of study leading to the award of a degree, and how the University controls their choice, assesses their work and awards degrees. Section 3.7 describes the sample of students whose academic records are analysed in chapters 4 to 6 and summarises the information obtained for each student.

3.2 The Modular Degree Programme at Oxford Brookes University

The Oxford Polytechnic Modular Degree Programme first admitted students in September 1973 and was then a course in science with seven fields of study, the first multi-disciplinary modular course to be validated by the CNAA (Watson, 1989). The former polytechnic, Oxford Brookes University, now offers a much larger modular programme in which over forty single subject honours degrees are available. In addition, over 1,000 joint honours degrees are available since most subjects can be studied in combination with one of more than 40 others. Subjects named in single or joint honours degrees are referred to within the Modular Programme as 'fields'. 'Double fields' lead to single subject degrees and joint honours degrees are based on two 'single fields'. This chapter describes the general structure of the Modular Programme, giving the detailed regulations as they applied to students entering the Modular Programme in 1994.

The academic year is divided into three terms and the majority of modules are taught and assessed within one term. Modules are offered at two levels, described as 'basic' and 'advanced': with each module taught and assessed independently of all the other modules. Students study for a degree by following a programme made up of individual modules: in a typical term a full-time student takes three modules. Students' programmes are divided into two stages corresponding to the first and subsequent years of a three-year, full-time degree course. In the first year (stage 1) a student's programme consists entirely of basic modules. This year forms an important part of the student's degree but the marks achieved do not contribute to the calculation of degree classifications. A student who has satisfied the stage 1 requirements for their degree, progresses to stage 2 (the second and third years of their course) in which their programme consists of 'acceptable', mostly advanced, modules. Degree classifications are based on the marks achieved in some, but not all of the modules taken in stage 2.

3.3 Content of programmes

To be awarded a degree, a student must accumulate the required number of module credits by passing modules whose content 'adds up' to a valid degree. Although students choose which modules to take, the university controls the content of students' programmes: by making some modules compulsory or required, by controlling access to modules through a system of pre-requisites and exemptions, by restricting students' choice of modules in stage 2 and by requiring potential honours graduates to achieve two module credits for a final year project or dissertation.

Modules that are compulsory or required for a particular degree define the core curriculum for that award. These modules must be included in a student's programme and to meet the requirements of their chosen degree, the student must pass all the compulsory modules and achieve a mark of at least 30% in each required module. Pre-requisites and exemptions are concerned with which students may enter a module and are expressed in terms of other modules within the Modular Degree Programme or students' entry qualifications. Exemptions are used to make a basic module compulsory for some students but not for others; for example students who have covered some background to their degree in their A-level studies may be exempted from having to take a module which is essential and therefore compulsory for those who have not. Restrictions prevent students from obtaining multiple credit for the same attainment, by stating that a credit for one module cannot be counted in addition to a credit for another whose syllabus overlaps the first. Pre-requisites help to create depth in students' programmes as students who wish to take a module that has one or more pre-requisites must first achieve a specified level of performance in the pre-requisite modules. This allows the creation of specialist pathways or strings of modules in which students can acquire a deep knowledge of a particular area.

In stage 2, the second and third years of a three-year degree course, students' choices are restricted by the need to take at least 16 modules from the list(s) of modules defined as 'acceptable' for their field(s). The acceptable modules for a field are usually advanced but may include one or two modules that are 'acceptable basics'. Acceptable basic modules are used to provide an interdisciplinary element in a degree course, consequently modules which are merely 'basic' in one field may be defined as 'acceptable' in another. Results achieved in such a module will contribute to the degree class of a student in the second field but not in the first.

The award of an honours degree depends, inter alia, on the student completing a substantial piece of independent work, in the form of a final year project or dissertation. Full-time students work on their final year project during the first two terms of their final year, while continuing to take other modules, and submit their work for assessment at the start of the third term. If passed, this generates two module credits. Without these two module credits, students can only achieve an ordinary degree.

3.4 Planning a programme

Students are responsible for choosing, registering and updating their own programmes using the information and support systems available. Stage 1 (first year) programmes are first planned when the student enters the Modular Degree Programme and stage 2 (second and third year) programmes are first planned in the second term of the first year. Changes to the programme can be made at any time.

Students must plan a programme that will allow them to meet the requirements for the award of an honours degree. For students entering the Programme in 1994, these were as follows:

* pass all compulsory modules
* achieve at least 30% in all required modules
* meet the pre-requisite requirements for all the modules in their programme
* achieve two module credits for a final year dissertation or project
* pass 9 basic modules in stage 1

* pass 27 modules in total either within 9 terms or, if over an extended period, within no more than 21 modules attempted in stage 2

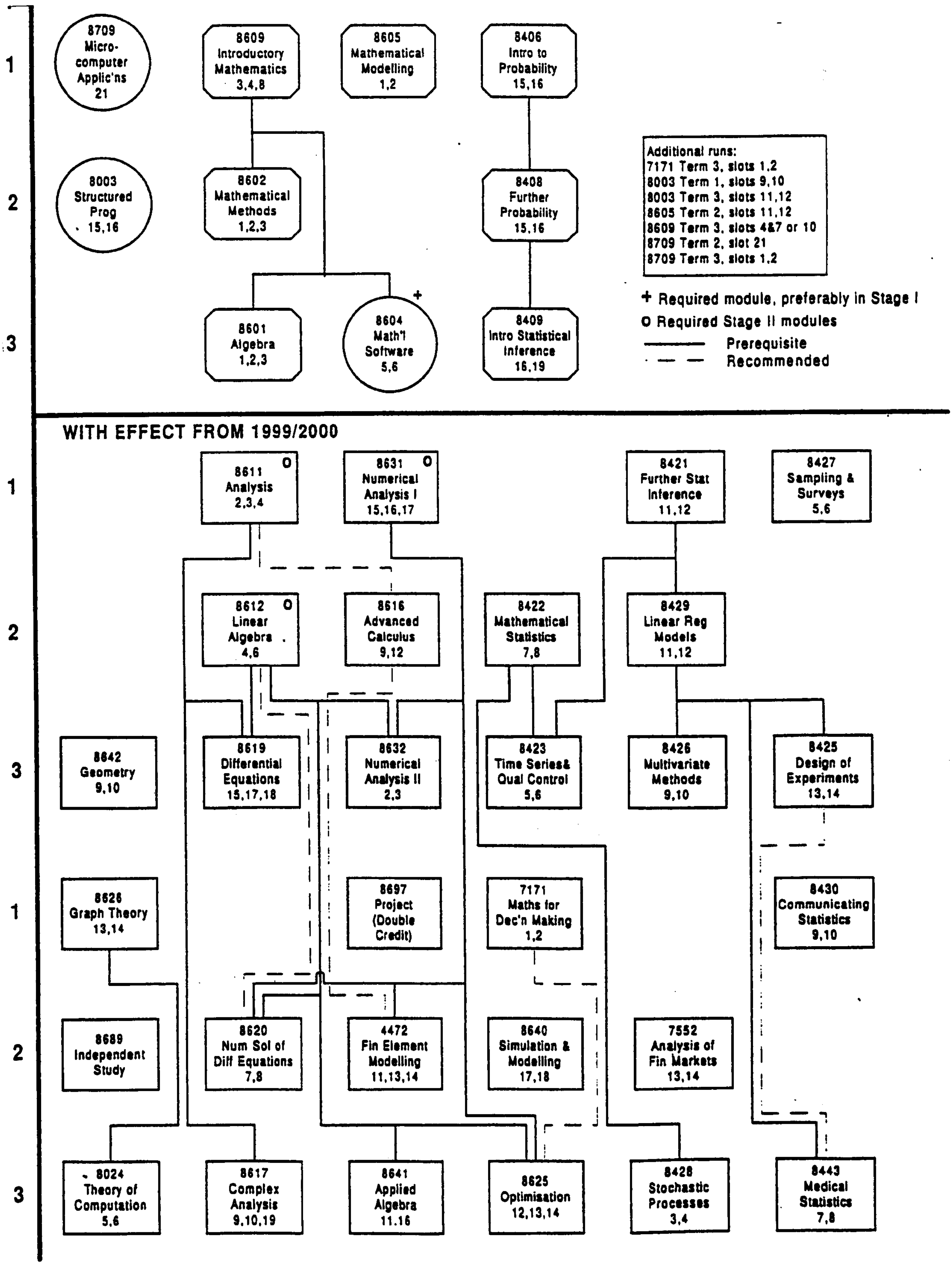
* pass at least 16 acceptable modules including, for combined honours, a minimum of 7 in each field, with at least six credits in acceptable modules other than the final year dissertation (MDP, 1994).

Students are provided with the handbook(s) for their chosen field(s) when they enter the Modular Programme (MDP, 1994). These handbooks include field diagrams showing the compulsory, required and acceptable modules and the relationships between them. An example is given in Figure 3.1, which shows the field diagram for the B.Sc. in Mathematical Sciences, a single subject degree. Here, modules serving different purposes are represented by different symbols: compulsory modules are represented by lozenge shapes, acceptable modules by rectangles, recommended basic modules by circles. (Recommendations are in the spirit of 'helpful suggestions' which students are free to follow or ignore as they wish.) The diagram shows when, in the academic year, each module is taught and the numbers printed beneath each module title indicate when the module is time-tabled, so that students can check for clashes. Acceptable modules are shown in the part of the diagram that represents the second and third years. Modules and their pre-requisites are joined by solid lines, while dotted lines show where students are recommended to take one module before another. Not all of the modules in a student's programme must be chosen from the field diagram(s); later, it will be seen that some modules may be chosen more widely. Apart from the field diagrams and lists of compulsory, required and acceptable modules, the field handbooks provide detailed descriptions of each module in the diagram. These describe the modules' pre-requisites and exemptions, content, intended learning outcomes, teaching methods (that is, lectures/seminars/laboratory sessions etc.) and assessment methods.

Constructing a programme is simpler for students who are studying for a degree in one subject, as they have only one set of requirements to meet; combined honours students need to work from two diagrams, one for each field.

The first step in planning a first year programme is to identify all the modules that are compulsory or required for the student's field(s) and their pre-requisites. All of these must be included. Additional modules may be chosen from the basic modules recommended for a student's field(s) or from basic modules in any subject, as long as timetable clashes are avoided and the student can meet the appropriate pre-requisite requirements.

Figure 3.1 Field diagram, Mathematical Sciences



Planning a stage 2 programme also begins with the selection of compulsory and required modules and their pre-requisites, including a final year dissertation or project. Students studying for a combined honours degree must decide whether to carry out a project in one of their fields (a 'double' project) or both (an interdisciplinary project). In a combined honours degree, a 'double' project will contribute two credits to the student's overall module total, but only one credit to the total for the appropriate field, while an 'interdisciplinary' project generates one module credit in each of the two fields. In addition to the compulsory and required modules listed above, acceptable modules must be added until the programme includes at least 16 acceptable modules (leading to at least 7 credits in each field, for combined honours). After meeting these requirements, students entering the Modular Programme in 1994 could choose additional modules from any of the following categories (assuming that the pre-requisites were met): modules which are listed as acceptable for the student's field(s), advanced modules in any subject, and up to two 'non-acceptable' basic modules. Marks achieved in modules in the first two of these categories were considered in the calculation of degree classifications, marks achieved in non-acceptable basic modules were not.

3.5 Assessment within modules

Both examinations and coursework are used to assess students in the Modular Degree Programme and within an individual module, one or both of these methods may be used. The method of assessment used in each module is one of many items of information available to students planning their programmes. With the exception of final year projects or dissertations, assessment is carried out during a module or at the end of term. Setting the assessment for the module is the responsibility of the module leader. In modules that use an examination to assess students, a fresh paper is written on each occasion the module runs. Each examination paper is reviewed before the start of the module, by an internal assessor and the external examiner for the relevant subject. In modules that rely entirely on coursework assessment, the internal assessor and external examiner review the assignments to be set. Module leaders are responsible for assessing students' work, gathering together medical certificates and other documentation referring to the module's intake of students, producing an unmoderated mark sheet and writing a report on the results for the external examiner. Double marking and the marking of anonymous scripts are not routinely required by the institution. Students who complete the assessment tasks are awarded a mark out of 100, with a corresponding grade. Instructions for module leaders remind them to use the whole range of marks and of the meaning of each grade. When all the work has been marked, samples of students' work, marksheets and the module leader's report on each module are sent to the appropriate external examiner. Subsequently the results achieved in each module are confirmed or moderated by the relevant committee of subject examiners, with the external examiner acting as ultimate arbiter.

Two members of staff drawn from the field(s) to which the project applies assess final year projects or dissertations. These assessors must agree a mark between them. As the project is a 'double' module, students who pass receive two module credits and the agreed mark carries twice the weight of the mark for a single module in calculations of degree class. Interdisciplinary projects come under the aegis of the external examiner for one of the two fields, who receives a project report recording the views of both internal assessors and has access to all relevant projects.

In fields that recruit large numbers of students, the external examiner may decide to inspect a sample of projects each year rather than all those submitted.

The pass mark and definitions of grades are the same for all modules: the pass mark is 40% and grades are defined as follows:

Below 40%	= F (failure)
40% and over	= C
50% and over	= B
60% and over	= B+
70% and over	= A

Additional grades are listed below. Those marked with an asterisk are temporary codes.

MC*

Awarded in special circumstances, this indicates that the exam committee wishes to moderate a student's mark, but will do so at a later date, when there may be more evidence of how the student performs.

MS Medical satisfactory

An unqualified pass grade awarded when the student has failed but this is judged to be the result of certified illness or other special circumstances.

RE*, RC*, RB* Resit grades

The second letter indicates the assessment to be repeated: E = examination, C = coursework, B = both. Resits are normally awarded only when a student fails a basic or compulsory module, but achieves a mark of at least 20%. These grades are temporary, if a student does not resit the module, their original mark is re-graded as an F. After taking the resit, the mark is changed to the mark achieved, and grade F, in the event of failure or 40, grade P in the event of a pass. Students may resit only one failed module from each term, by taking a resit exam and/or submitting resit coursework in week 3 of the term following their failure, or in September if the failure occurred in the summer term. If a resit is not taken, a student's original mark is re-graded as an 'F' indicating failure. Students who do not achieve sufficient marks in a required or compulsory module at their first attempt or in a resit, must repeat the module.

MR* Medical resit

If a student is unable to complete the assessed work for a module as the result of illness or other special circumstances, they may be offered a medical resit. The conditions outlined above apply to medical resits, but the student's final mark is not restricted to 40%.

39 F/RE*/RC*

A mark of 39 is reserved for 'technical failures'. Modules which use both coursework and examination may have, in addition to the 40% pass mark, thresholds which must be achieved in one or both components. Where this is the case, students who achieve an overall mark of 40 or more, but have not met the requirement for one of the components, are allocated a mark of 39. Students with an initial overall mark of 39, who are not 'technical failures', have their mark moderated to either 38 or 40.

F0

A mark of F0 is recorded when a student registers a module within their programme, but submits no work for assessment. If a student registers for module, decides not to take it but does not delete it, then an F0 appears on the student's record and is counted as a failure within the Modular Programme's regulations. F0's have been excluded from the sample to be analysed here along with other marks of 5 or below.

These are likely to represent modules in which students attempted only part of the assessment.

CR External credit.

This grade indicates that a student has been credited with a module pass on the basis of study in another institution.

3.6 Confirmation of results and feedback to students

The Modular Degree Programme at Oxford Brookes has a two-tier system of examination boards, which is typical of modular higher education. Three times a year, at the end of each assessment period, subject examination committees review the outcomes of assessment in each module for which they are responsible. The progress of students in the field or group of fields covered by the examination committee is also reviewed. The committee does not have the power to make decisions about any student's progression or graduation: these are the responsibility of the second tier of examiners, but the subject examination committee may pass on their views or comments to the second tier of examiners. The second tier of examiners includes the chief external examiners who consider special cases and degree awards. When both tiers of examiners have confirmed or moderated the results, students' academic records are updated and copies are distributed to personal tutors and the students themselves. Error messages are generated if the student's programme or achievements do not meet the requirements of their degree or if the programme includes a module with no provision for the student to achieve the necessary pre-requisites. Other error messages may warn the student that, perhaps as the result of failure, they need to add modules to their programme. Programme errors that are not corrected will prevent the student from progressing or graduating as intended. When programme errors are identified, an additional message is printed, advising the student to consult their personal tutor, whose name is included in the record.

Academic records have the same format throughout a student's degree and show the student's chosen programme, the marks and grades achieved in each module taken, the number of modules passed at each level and the average marks achieved. The student's average is calculated from the marks achieved in modules that the student has passed and the student's status determines which module marks are used. In stage 1, the average is calculated from all modules passed, in stage 2, the average is calculated using only the marks achieved in acceptable or advanced modules passed. For students graduating in 1996/7, the average of the best 16 marks achieved in acceptable modules was used to decide the classification of their degree if they have met the criteria for an honours degree. Figure 3.2 shows an example of a student record. As this student has graduated, their record also shows their degree classification.

Figure 3.2 Example of a student record

```
        STUDENT'S NAME                                         TITLE

       STUENT NUMBER                                                 ADDRESS

            31-MAY-75
            26-SEP-94
            04-JUL-97             04-JUL-97
            COMPLETED COURSE OK
        LEA: 2903
0806 BRITISH CITIZEN                     02 ENGLISH OR WELSH LEA AWARD
        MC10  MOD DEGREE/DIP HE/CERT HE            MC10B1 STAGE II HONS MODULAR                 FULL-TIME
        BA Hons Modular                            PERSONAL TUTOR
        EN    ENGLISH STUDIES            Field Chair :
        PB    PUBLISHING                 Field chair:

                   Entry qualifications: AENE B    AGG  C    ATH  C    SGS  B

SEP-94 STAGE I MODULAR COURSE         APR-96                                                      STAGE II HONS MODU
      M01501 B    58 B    HUMAN BE    EN    M02366 A    60 B+   CONT LIT                          Taken and Credited
en    M02305 B    54 B    LANG/LIT    PB    M05232 A    65 B+   MARKETIN                          Modules
pb    M05200 B    58 B    KEY CONC    PB    M05251 A    60 B+   EDIT MAN
                                                                                                  Passed       28
JAN-95                                SEP-96
pb    M05207 B    69 B+   MAKING B    EN    M02313 A    61 B+   MED DRAM                          Basic         9
                                      EN    M02353 A    68 B+   MOD LIT:                          Total Basic   9
APR-95                                EN    M02367 A    62 B+   CONTEMP                           Acceptable
en    M02306 B 2  61 B+   TEXT/PRO    PB    M05219 A    69 B+   DESK-TOP                            Basic       0
pb    M05208 B    64 B+   PUBLISH/                                                                  Advanced   19
      M05209 B    64 B+   CONTEMP     JAN-97                                                        Total      19
      M08709 B    62 B+   MICRO AP    PB    M05252 A    68 B+   EDIT TEC                              EN       12
                                                                                                      PB        7
SEP-95 STAGE II HONS MODULAR          APR-97                                                          either    0
EN    M02343 A    62 B+   VIC LIT:    EN    M02318 A    65 B+   CHAUCER                           Other
EN    M02361 A    59 B    CONT LIT    EN    M02355 A 2  70 A    MOD LIT:                          Advanced      0
                                      EN    M02392 A 2  58 B    CRITICAL
JAN-96                                                                                            Average      64.2%
EN    M02388 A    60 B+   EN IND S                                                                over best    15
PB    M05211 A 2  62 B+   PUBLISH
PB    M05218 A    59 B    ELECT PU                                                                Acceptable
                                                                                                  Taken        19
                                                                                                  BA Hons Modular
                                                                                                  upper second class
                                                                                                  Awarded
                                                                                                  17-JUL-1997
```

3.7 Classification of honours degrees

For students who qualified for an honours degree in 1997, degree classifications were based on the average of the best 16 marks achieved in acceptable and/or advanced modules, calculated to one decimal place. In this calculation, as in the calculation of module credits, marks achieved in double modules carry double the weight of marks achieved in single modules. The thresholds for each degree class are defined as follows:

70% and over	1st
60% and over	2:i
50% and over	2:ii
40% and over	3rd

This system classifies honours degrees 'automatically', but some degree classifications may be raised to a higher class through the application of the university's borderline procedures. For students graduating in 1997, a degree in the higher class would be considered only if three criteria were met: the mean mark for the student's best sixteen modules fell just below a boundary between two classes; a majority of the eight most recent of the student's best sixteen marks were in the higher class; the relevant subject examination committee(s) were in favour of awarding the higher class. To meet the first criterion, a student's average had to be within 0.5 marks of the lower boundary for lower second class honours or within 1 mark of the lower boundaries for upper second or first class honours. If all three borderline criteria are met, the decision is referred to the second tier, of chief examiners, who would normally decide to raise the student's degree to the higher class.

3.8 Ordinary degrees

A student graduating in 1997 was awarded an ordinary degree if, having successfully completed stage 1, they achieved 16 credits in acceptable and advanced modules in stage 2. These may or may not include credits for a dissertation. Regulations limit the number of modules that can be taken and the length of time a student may take to achieve an honours degree. All students enrol for honours degrees but may adjust their award aim during their programme if they wish. Students aiming for honours may be awarded an ordinary degree if they qualify for one and their record shows that they cannot achieve an honours degree within the regulations.

3.9 Selection of the sample

The cohort studied here consists of all those students who enrolled on the Modular Degree Programme at Oxford Brookes University in September 1994 and graduated in the summer of 1997, excluding students whose degrees involved a compulsory period of assessed practice. A total of 496 students met these conditions. As a result of this selection method, several groups of students who enrolled on the Modular Degree Course during the same year have been excluded: students entering the course at other points in the academic year, students who transferred to Oxford Brookes from other universities, part-time students, students who withdrew or delayed their graduation, and students who enrolled on degree courses of 4 years duration. Students who enrolled to study a language have been excluded, as they are required to spend a year abroad. Students who enrolled on sandwich degree courses were excluded as their courses involve a compulsory year in a work placement and are following programmes that have a four-year duration. Students enrolled on degree courses in occupational therapy, nursing or health care studies were also excluded since their degree courses also involve a substantial, compulsory element of practice and are generally four years in length. These exclusions were made in order to obtain a cohort of entrants who had completed degree courses with the same duration, mode of study and course structure. As a result of these exclusions, care is needed in applying conclusions based on the chosen sample to other contexts. This issue will be discussed in chapter 7.

Information relating to the students' background and previous education, their programmes of study, academic achievement and the learning and assessment environment experienced in each module taken was extracted from Oxford Brookes' Computer Student Management System (CSMS). Each record in the data file represents one module entry, so that each student contributes approximately 28 responses. Note that double modules are represented by two records, one for each term in which the student participated in the module, and otherwise identical. In the data file, marks are associated with the term in which the student worked on the module rather than the term in which their final marks were confirmed. The variables available are listed in Table 3.1.

Table 3.1 List of variables extracted from students' records

For each student:

Age
Sex
Entry qualifications
Domicile: UK students by LEA, overseas students by country
Social class of parent's occupation

For each student's programme:

Field(s) studied
Number of modules taken
Number of module passes:
- in basic modules
- in total
- in acceptable modules
- in acceptable basic modules
- acceptable modules in each field, in both fields, in other fields

Number of modules taken in each term
Average (for degree classification)
Degree award and class

For each module within a programme:

Module number (id)
Term taken
Level (basic/advanced)
Number of credits (double/single)
Mark achieved
Number of students enrolled
Percentage examination assessment
Subject group responsible for teaching and assessment
Type of module (final year project/other)

3.9.1 Distribution of sample by age, sex and parent's social class

Each student's age, sex, domicile, parent's occupation and previous education are recorded in the processes of application and enrolment. Table 3.2 shows the distribution of the cohort by age and sex. A majority of the students are women and 23.2% were 'mature', that is aged 21 or over, when they enrolled on the Modular Degree Programme. The proportion of mature students is slightly higher amongst women students (24.1% compared to 21.8% amongst male students). 85.5% of the sample had home addresses in the United Kingdom.

Table 3.2 Distribution of Sample by Age and Sex

AGE	MALE		FEMALE		TOTAL	
On September 1st 1994	n	%	n	%	n	%
20 and under	161	78.2	220	75.9	381	76.8
21-29	27	13.1	34	11.7	61	12.3
31-39	9	4.4	24	8.3	33	6.7
40 and over	9	4.4	12	4.1	21	4.2
TOTAL	206	41.5	290	58.7	496	100

During enrolment, students are asked to provide information on parents' occupations. The students' answers are coded by the university's administrative staff, and for the purpose of this study, the occupation codes have been used to derive social class categories. Table 3.3 shows the distribution of the cohort by social class; 61.7% of the sample had a parent whose occupation was classified as a professional or managerial, and 20.1% a parent in a manual occupation, 1.8 % of students are unclassified, as the result of having failed to provide any relevant information.

Table 3.3 Distribution of Sample by Social Class of Parent's Occupation

Social Class, parent's occupation	n	%
Not stated	9	1.8
I	83	16.7
II	223	45.0
III non-manual	71	14.3
III manual	75	15.1
IV	21	4.2
V	4	0.8
Armed forces	10	2.0
Total	496	

3.9.2 Distribution of Sample by Measures of Prior Achievement

Students' qualifications are recorded when they enter the Modular Degree Programme: Table 3.4 shows the types of qualifications listed on students' records.

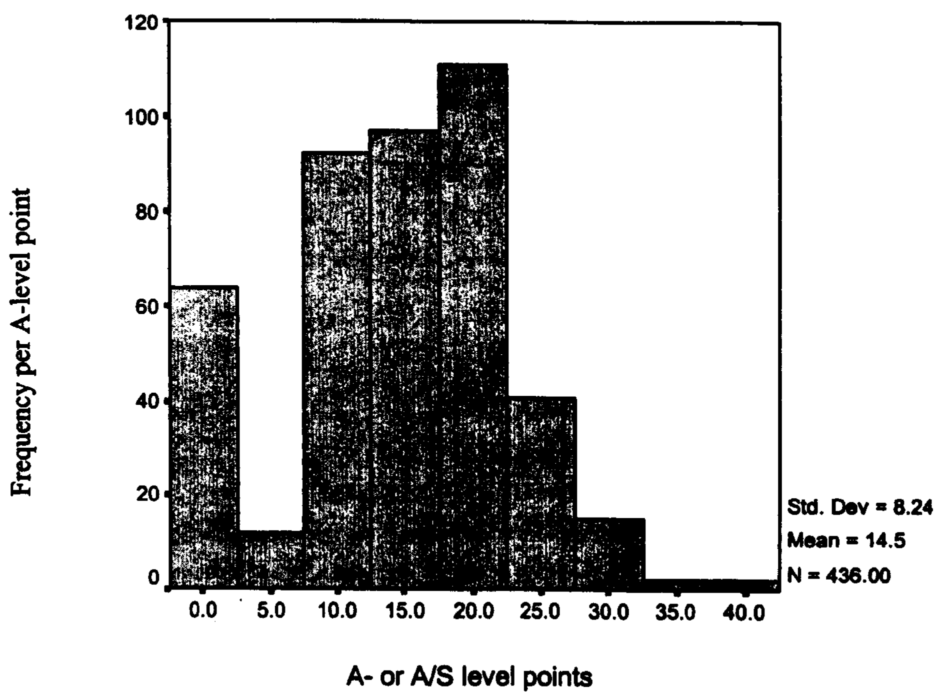
Table 3.4 Distribution of Sample by Type of Entry Qualification

Qualification	n	%
A or A/S levels	436	87.9
GCSE/O-level	14	2.8
Sub-degree qualification	21	4.2
Professional or intermediate qualification	9	1.8
Other, not recorded	16	3.2
Total	496	

In this table, students with more than one kind of qualification have been allocated to a single category by giving priority to higher level qualifications, rather than most recently obtained. This means that, for example, students with one A-level pass who have entered higher education as the result of taking an Access course are included in the first category of the table.

More detailed information records the subjects taken and the grades achieved by students with A or A/S level passes: these are used to calculate a score that is used as a general measure of achievement. The score is calculated by awarding 10 points for each A-grade, 8 points for each B-grade, 6 points for each C-grade, 4 for each D-grade and 2 for each E-grade achieved at A-level. A/S levels are scored in the same way, but each grade is awarded half the points that would be given for the same grade at A-level. For each student, the points awarded for each A or A/S level pass are then added together. Figure 3.3 shows the distribution of these scores for students identified as having A or A-S level qualifications; due to missing data, many students appear to have zero scores.

Figure 3.3 Histogram of A and A/S level scores



For the 380 students (76.6 % of the sample) with non-zero scores, the mean A- and A/S-level scores for students belonging to different groups are shown in Table 3.5. This table shows that women and students entering the course below the age of 21 had somewhat higher mean scores than other students, with only small variations in scores by social class.

Table 3.5 Mean A-level points for students in different groups

	Mean	Standard deviation	n
SEX			
Male	15.7	5.79	154
Female	17.5	6.54	226
AGE			
20 and under	17.1	6.07	341
21-29	14.4	7.47	31
30-39	16.5	7.72	4
40 and over	8.5	7.55	4
SOCIAL CLASS OF PARENTS			
I	17.0	7.26	65
II	17.3	6.31	186
IIINM	16.3	5.48	522
III M	15.6	5.98	44
IV	16.2	5.53	17
V	17.3	5.03	3
Other, not known			13

Students' prior achievements are of particular interest in the analysis of their achievement as undergraduates, as when comparisons are made between the achievements of sub-groups of students, controlling for previous attainment is desirable. Problems arise when attempting to control for the prior attainment of undergraduates since more than one kind of qualification is recognised in the selection of students for admission to higher education and so no common measure of attainment is available for all students. A second problem is that the recording of entry qualifications by universities tends to reflect the requirements of the Higher Education Statistics Agency (HESA) which, for 1994 entrants did not require detailed information such as A-level grades. Consequently, even where students had entered higher education on the basis of their A-level passes, the information recorded by the university may be incomplete. In the sample studied here, of the 436 students described in Table 3.4 as having A or A/S-level qualifications, 375 (86.0%) had A-level grades entered on their student records. A further five students, with 'sub-degree' qualifications on entry to the course, also had A-level grades entered in their record. This means that a score of zero may indicate either a student who has no A or A/S level passes or a missing response and a non-zero score represents only a part of a student's attainment at the start of their degree programme, if the student has other types of qualifications. In practice, the A-level grades recorded by the university provide a poor measure of the prior attainment of this cohort.

Although there are now a variety of routes into higher education, entry requirements are commonly stated in terms of A-level grades. The Oxford Brookes' prospectus for students applying for entry in September 1994 states the A-level entry grades which students need to gain a place on their chosen degree course. As some subjects are more popular than others, even within an institution, the A-level grades required may vary from one subject to another. For each student's award, the score equivalent to these entry grades was calculated and the distribution of these scores is shown in Table 3.6.

Table 3.6 Distribution of students' scores corresponding to entry grades for chosen degree award

Score	Frequency	%
8	59	11.9
9	6	1.2
10	15	3.0
11	5	1.0
12	62	12.5
13	22	4.4
14	36	7.3
15	25	5.0
16	23	4.6
17	37	7.5
18	62	12.5
19	33	6.7
20	26	5.2
21	20	4.0
22	30	6.0
23	7	1.4
24	28	5.6
Total	496	

When these scores are compared to the students' own recorded A and A/S-level scores, 67 students (6.9% of the sample and 17.9% of those described in Table 3.4 as having A or A/S level qualifications and with recorded grades) have non-zero scores which are more than 2 points below the entry grades specified in the prospectus. There are several possible explanations for these discrepancies: the recorded A-level grades may be inaccurate, incomplete or represent only a fraction of a student's portfolio of qualifications. Some students may have changed their award aim during the course of their degree to one associated with higher entry grades than were required by the student's first choice, others may have been admitted through the clearing process on the basis of lower grades than those published in the prospectus.

Given the uncertain meanings of recorded entry qualifications and A/A/S level scores, there would be difficulties in interpreting the findings of any analyses in which these measures were used as independent variables. Analyses of student achievement presented in later chapters will therefore avoid these measures, but will use the entry grade scores for each student's degree award as a potential independent variable. Strictly speaking, these scores measure the expectations of the university with respect to students following a particular course, rather than the individual student's achievement before entry and this will need to be considered when the analyses are interpreted.

3.9.3 Subjects of students' degree awards

324 students (65.3% of the sample) were awarded combined honours degrees: this is higher than the proportion of all students graduating from Oxford Brookes with combined degrees in 1997 as the exclusions listed in section 3.8 remove a disproportionate number of students taking single-subject degrees. Table 3.7 shows the numbers of students in the sample following programmes administered by the schools or departments listed. Note that the Law department has been shown separately as this department moved from the School of Business to the School of

Social Sciences (which then became the School of Social Science and Law) during the period in which the sample were enrolled. Two hundred and five students appear twice in this table, having taken degrees combining fields administered by different schools within the university. 119 students were awarded degrees combining fields administered by different schools or departments.

Table 3.7 Distribution of sample by school/department responsible for field(s) studied

School/department	Men n	%	Women n	%	Total n	%
Biology and Molecular Sciences	23	11.2	64	22.1	87	17.5
Planning	30	14.6	20	6.9	50	10.1
Business	68	33.0	79	27.2	147	29.6
Law	20	9.7	23	7.9	43	8.7
Social sciences	38	18.4	72	24.8	110	22.2
Construction and earth sciences	26	12.6	11	3.8	37	7.5
Engineering	3	1.5	1	0.3	4	0.8
Visual arts, music and publishing	18	8.7	32	11.0	50	10.1
Computing and mathematical sciences	26	12.6	18	6.2	44	8.9
Education	2	1.0	16	5.5	18	3.6
Humanities	27	13.1	65	22.4	92	18.5
Hotel and restaurant management	4	1.9	8	2.8	12	2.4
Combined studies	1	0.5	6	2.1	7	1.4
Total	206		290		496	

Table 3.7 shows that female students were more likely than men to have obtained degrees in fields managed by the Schools of Biology and Molecular Sciences, Social Sciences, Education and Humanities and male students were more

likely than women to have obtained degrees in fields managed by the Schools of Planning, Business, Construction and Earth Sciences and Computing and Mathematical Sciences. The specific fields taken by students in the sample are shown in Tables 3.8 and 3.9, showing the subjects studied by students graduating with single subject and combined degrees respectively. Note that students who obtained combined degrees appear twice in Table 3.9.

Table 3.8 Distribution of sample by school or department: single subject degrees

Fields	n	%
Nutrition and food science	7	4.1
Planning studies	33	19.2
Business administration and management	21	12.2
Human biology	17	9.9
Law LLB	28	16.3
Psychology major	4	2.3
Cell and molecular biology	6	3.5
Environmental biology	7	4.1
Applied Geology	13	7.6
Applied physics	3	1.7
Cartography major	5	2.9
Fine art	13	7.6
Geological sciences	5	2.9
Mathematical sciences	4	1.2
Technology management	1	0.6
Curriculum studies/arts	1	0.6
Curriculum studies/history	2	2.3
Curriculum studies/geography	1	0.6
English studies	1	0.6
All students with single subject degrees	172	

Table 3.9 Distribution of students by field: combined degrees

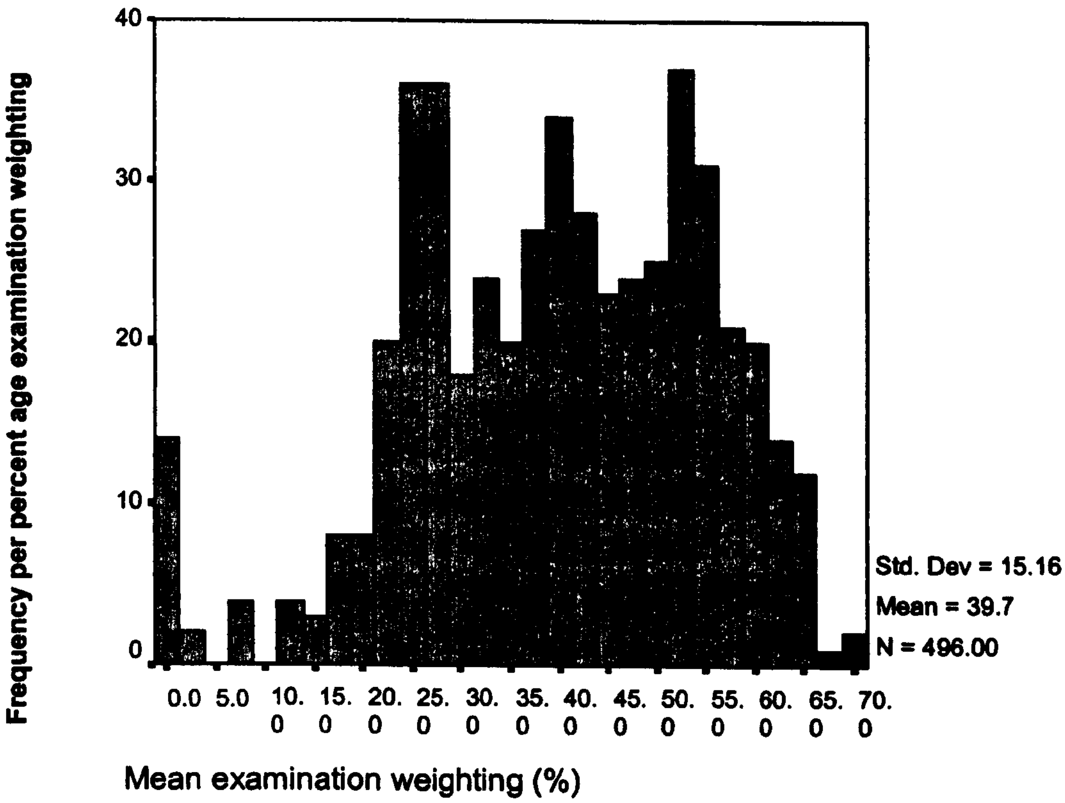
Field	n	%
History	35	5.4
Politics	17	2.6
English	48	7.4
History of art	35	5.4
Planning studies	15	2.3
Tourism	29	4.5
Retail management	16	2.5
Educational studies	14	2.2
Anthropology	35	5.4
Environmental sciences	19	2.9
Geology	8	1.2
Business administration and management	48	7.4
Geography	35	5.4
Accounting and finance	31	4.8
Music	2	0.3
Marketing management	27	4.2
Mathematics	12	1.9
Sociology	20	3.1
Cartography	7	1.1
Biology	16	2.5
Information systems	10	1.5
Publishing	30	4.6
Economics	26	4.0
Intelligent systems	4	0.7
Fine art	6	0.9
Statistics	2	0.3
Computing	14	2.1
Chemistry	2	0.3
Psychology	19	2.9
Law	15	0.8
Hospitality management	12	1.9
Exercise and health	7	1.1
Food science and nutrition	5	0.8
Applied physics	8	1.2
Microelectronics	3	0.5
Computing mathematics	2	0.3
Combined studies	6	0.9
Computer systems	1	0.2
Physics	1	0.2
Environmental biology	5	0.8
Geotechnics	1	0.2
Total single fields	648	

3.9.4 Students' programmes

On average, students had taken and passed more than the minimum number of 27 modules required for the award of an honours degree. On average, students had taken, 29.97 modules (sd = 1.77) and passed 28.0 modules in total (sd = 1.90), including 9.73 passes in basic modules (sd = 1.01).

In total, the students in the sample had made 14,371 module entries and marks were recorded for 14,315 of these entries (99.6 % of total). The remainder are either module credits obtained outside the Modular Degree Programme or for which individual marks are not recorded (for example MS grades).

Figure 3.4 Mean examination weighting experienced by each student

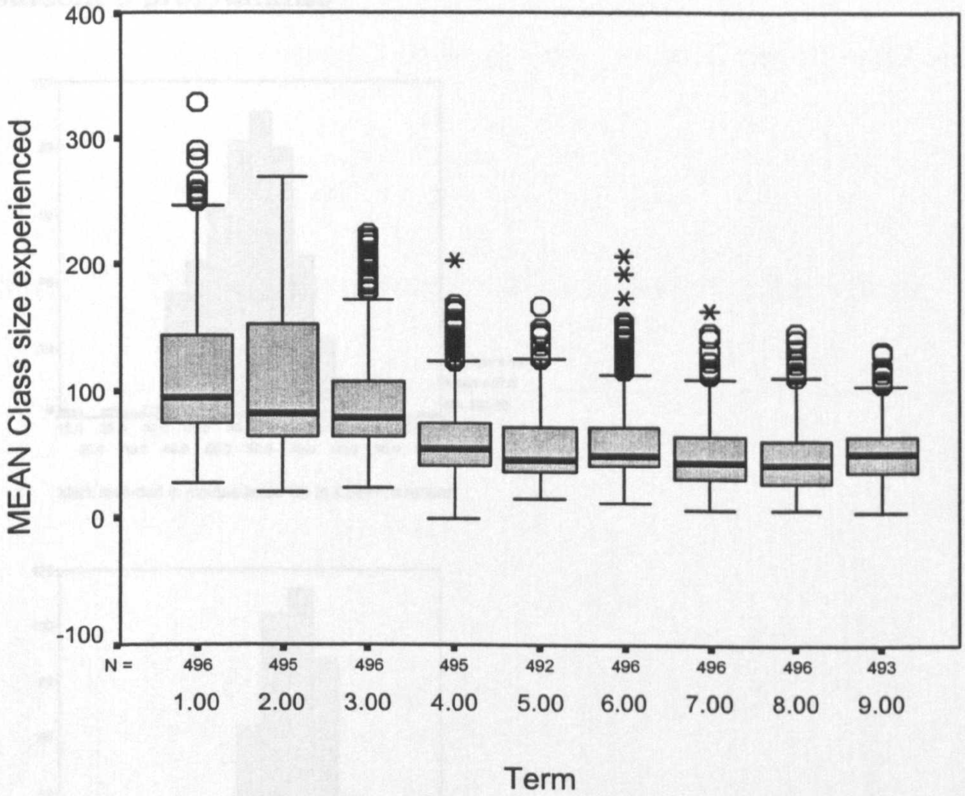


Modules provide students with a variety of experiences: each module has its own syllabus and teaching style, and modules also vary in terms of assessment methods, class size and student intake. These module characteristics help to shape a student's experience of the Modular Programme and may also contribute to variations in their performance. On average, the assessment in 8% (sd = 8.02%) of the modules in students' programmes was based entirely on examinations and in 37% (sd = 19.6%) of modules the assessment was based entirely on coursework. In the remainder of modules both coursework and examination assessment were used. As the percentage of examination assessment is recorded for all the modules in a student's programme, the proportion of all marks available for work produced in examinations can be calculated for each student's programme. The distribution of these values is plotted in Figure 3.4.

On average, the proportion of marks available in examinations is 39.5% (sd=15.28%), but Figure 3.4 shows how much this proportion varies from programme to programme. If the use of coursework assessment in higher education has an impact on students' marks, some variation in the degree awards of the sample studied here may reflect the fact that coursework assessment plays a larger part in some degree programmes than others. Within students' programmes, the percentage of marks available for coursework and examination varies from module to module and this may contribute to the variation in a student's marks. The variation in a student's performance is of particular interest since, as discussed in the previous chapter, the system for calculation of degree class used by the Modular Degree programme at Oxford Brookes University favours students whose performance is variable. A measure of the variation in coursework and examination weightings experienced by students is the standard deviation of the percentage of marks available in examinations for the modules in a student's programme. These standard deviations ranged between 0 and 44. The first of these values indicates programmes consisting entirely of modules using the same coursework and examination weightings. The mean of these standard deviations is 30.1 (sd = 7.4), indicating that a typical student takes modules which use a wide variety of coursework and examination weightings.

The mean class size experienced by a student depends on which modules are included in their programme: in this sample, for an average student, the mean class size experienced is 72.1 students (sd = 30.03), but the mean class size experienced by individual students ranges from 25 to 165 students. The term 'class size' is used here to refer to the number of students enrolled on a particular module. Figure 3.5 shows how some of this variation occurs: this diagram shows the distribution of the mean class size experienced by students in the sample, in each term of the three year period of their degree. Students who spent a term abroad or were absent for other reasons do not contribute data for every term, hence the small variations in the sample size from term to term. The graph shows a tendency for the mean class sizes experienced by students to be considerably larger in the first year and to decrease over the period of nine terms. In each term, the distribution of mean class size experienced by individual students is positively skewed. The relationship between class size and time needs to be taken into account when students' performance in modules of different sizes is compared. The variety of class sizes experienced by individual students can be measured by calculating the standard deviation of the class sizes of modules in each student's programme: over the whole sample, the average for this quantity was 52.73 (sd = 32.99).

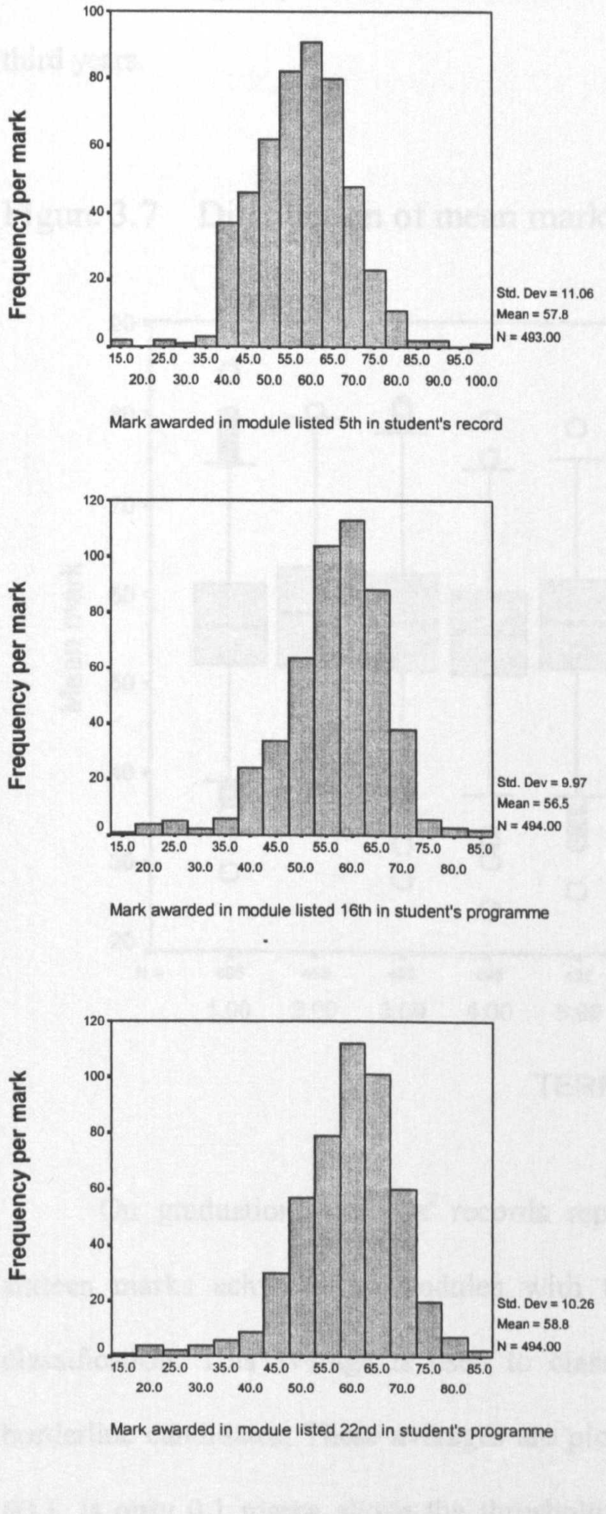
Figure 3.5 Average class size experienced by each student by term



3.9.5 Performance during degree programmes

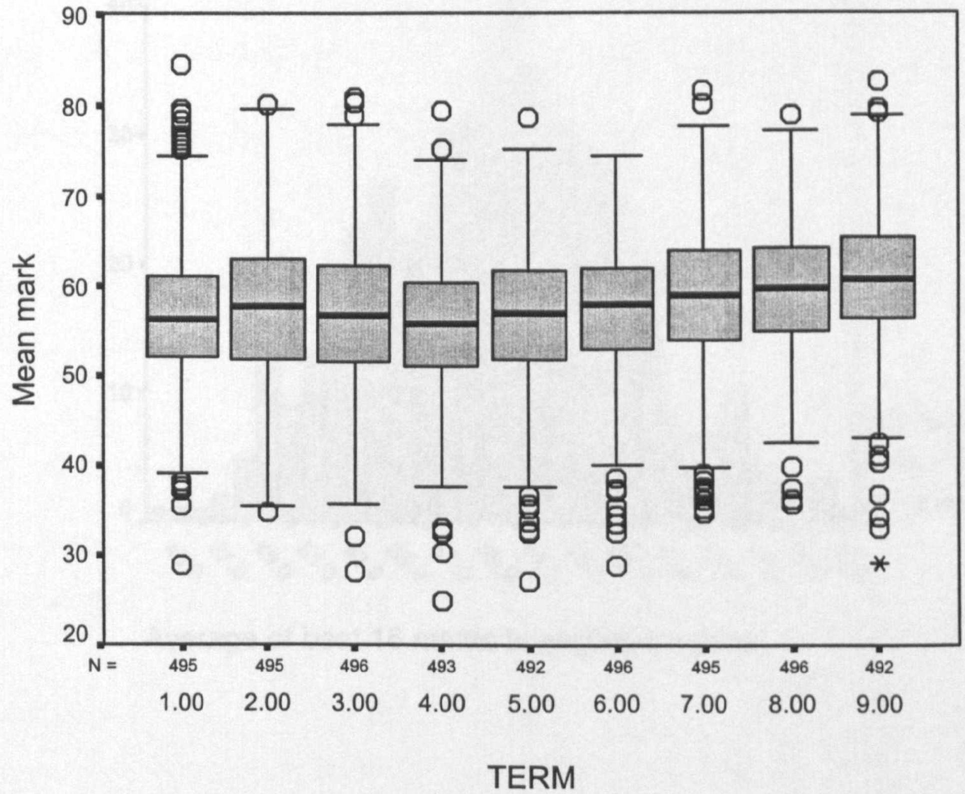
Figure 3.6 shows the distribution of marks achieved in modules listed 5th, 16th and 22nd in students' programmes and taken in the first, second and third years respectively of students' programmes. The mean marks achieved in these entries are 57.8, 56.5 and 58.8 marks respectively, with standard deviations of (11.05, 9.95 and 10.27 marks respectively). The mark distributions are all negatively skewed.

Figure 3.6 Distribution of marks for modules listed 5th, 16th and 22nd in student's programmes



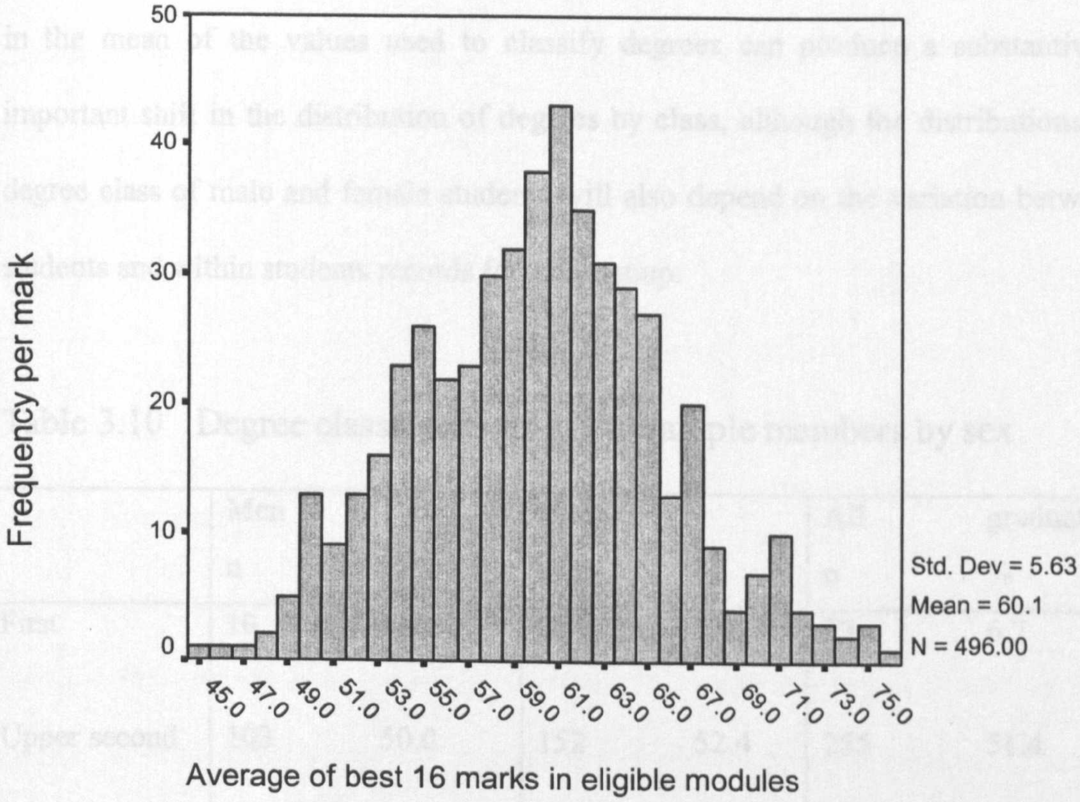
A simple view of how students' marks vary during the course of their degree is shown in Figure 3.7 which shows the distributions of students' mean marks in each term. A slight upward trend in the medians can be discerned in the second and third years.

Figure 3.7 Distribution of mean marks achieved in each term



On graduation, students' records report the average of the student's best sixteen marks achieved in modules with the potential to contribute to degree classification. This average is used to classify students' degrees and to identify borderline candidates. These averages are plotted in Figure 3.8. The mean value, of 60.1, is only 0.1 marks above the threshold between the upper and lower second classes. A small shift in this average could therefore produce a relatively large shift in the proportion of good degrees awarded.

Figure 3.8 Distribution of average of best 16 marks achieved in modules with potential to contribute to degree classification



After classification, the degrees awarded to the students in the sample are as shown in Table 3.10: in contrast to the figures provided by HESA (1998) for all first degrees awarded in 1996/7, in this cohort, women were more likely to achieve first class degrees than men. A potential explanation for this phenomenon is the higher proportion of mature students amongst female students. Table 3.11 shows that mature students were more likely to achieve first class or 'good' degrees than students who had entered the course at a more traditional age. Note that the percentages of women with first class or 'good' degrees were 3% and 5.4% higher respectively than the corresponding percentages of men with degrees in these classes. The averages used to classify degrees were 60.37 for women and 59.67 for men in the sample, a difference of only 0.7 marks. This suggests that a small change in the mean of the values used to classify degrees can produce a substantively important shift in the distribution of degrees by class, although the distributions by degree class of male and female students will also depend on the variation between students and within students records for each group.

Table 3.10 Degree classifications of the sample members by sex

	Men		Women		All	graduates
	n	%	n	%	n	%
First	10	4.9	23	7.9	33	6.7
Upper second	103	50.0	152	52.4	255	51.4
Lower second	82	39.8	102	35.2	184	37.1
Third	2	1.0	5	1.7	7	1.4
Ordinary	9	4.4	8	2.8	17	3.4
Total	206		290		496	

Table 3.11 Degree classifications of sample members by age

	20 and under		21 and over		All graduates	
	n	%	n	%	n	%
First	13	3.5	20	16.0	33	6.7
Upper second	198	53.4	57	45.6	255	51.4
Lower second	142	38.3	42	33.6	184	37.1
Third	5	1.3	2	1.6	7	1.4
Ordinary	13	3.5	1	0.8	17	3.4
Total	371		125		496	

3.10 Summary

This chapter has explained the framework within which the sample of students selected for this research accumulated academic credits leading to the award of a degree. Module marks entered in students' academic records are the results of students being assessed, module by module, term by term, within their degree programmes. The next chapter discusses how these marks can be analysed using models representing marks as arising from this hierarchical structure.

Several times in this chapter, relationships were seen between pairs of factors with the potential to influence student achievement: for example, a tendency was seen for median marks awarded to improve with time and cross-tabulations showed relationships between the subjects of students' degrees and gender. Relationships like these confirm the need, discussed in chapter 2, for multi-factorial analyses of performance.

This chapter has presented a number of tables and diagrams describing the sample of students in terms of age, sex and entry qualifications and degree subjects. The next chapter describes how the influence of student-level variables can be represented in a number of ways: as influencing students mean marks, as having an effect on the variability of an individual's performance or as influencing between student variation. Other information presented in this chapter showed that the module characteristics that are typical of one student's programme may present a different impression of a 'typical' module to that given by the modules in another student's programme. In particular, students following different programmes experience different weightings for coursework and examination assessment and different class sizes. The effects of these different experiences on students' marks will be explored in the analyses that follow in chapters 4 to 6.

The next chapter will propose a model for studying the effects of students' characteristics and the characteristics of their programmes on their recorded achievements.

Chapter 4

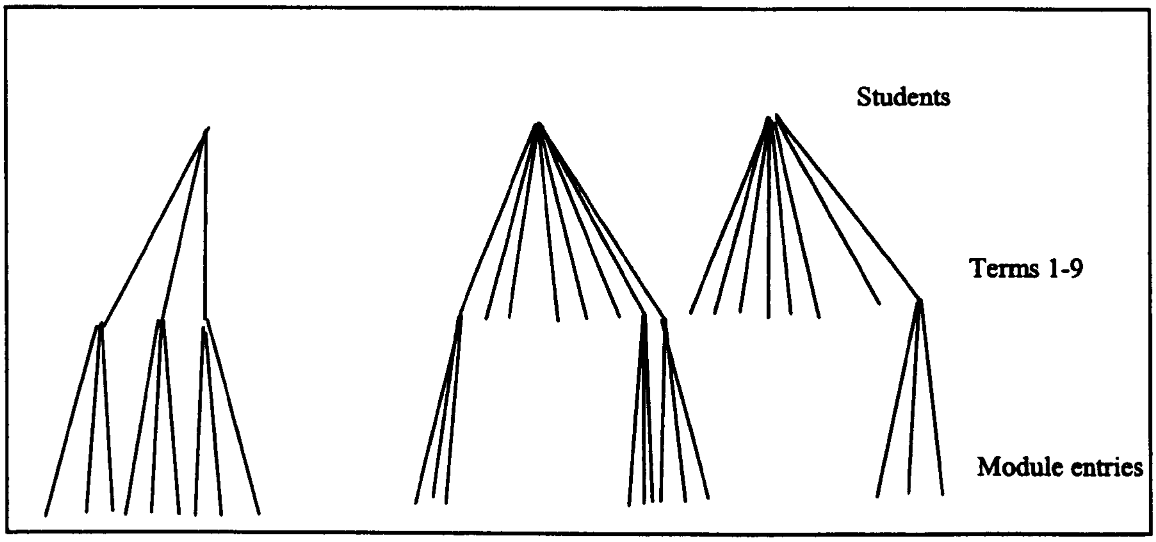
A Hierarchical Model of Student Achievement

4.1 Introduction

Chapter 3 explained how students enrolled in the Modular Degree Programme at Oxford Brookes University accumulate credits: over a period of three academic years, within up to nine terms, each student in the sample was assessed within each module taken. This chapter proposes a hierarchical structure for representing the assessment of students within their degree programmes and a series of models based on this structure is developed to meet the aims identified in chapter 1.

As each student took a number of modules, each student contributes a number of responses, recording the outcomes of assessments carried out at different times. The structure of the data can be represented as the three level hierarchy shown in Figure 4.1, where each student's marks are nested within terms and terms are nested within the student's programme. Assuming this structure, the data can be analysed by fitting models in which values of the dependent variable are the marks achieved in individual module entries (the first and lowest level units of analysis), grouped within terms (the second level units) which are themselves grouped within students' programmes (third and highest level units). The results of fitting these models to the data in order to investigate student achievement will be reported in section 4.5.

Figure 4.1 Three level hierarchical structure



Before considering models based on this structure, the next section discusses the form in which students' marks will be analysed.

4.2 Analysis of raw vs. standardised marks

The responses analysed are measures of students' achievements in a large number of different modules. Each of these modules has a unique syllabus and assesses students' performance in different coursework assignments and/or examinations; the marks awarded in each module are produced by the use of a different measuring instrument and therefore record achievements on a different scale. It is not unusual for a longitudinal study of educational achievement to involve measures of achievement recorded on different scales, since as pupils or students make progress, their attainment has to be measured in different ways. Students' achievements over a period of time can be studied more easily if a common scale is used to measure achievement on each occasion and when different instruments are used on each occasion, this can be achieved by standardising the responses, however, some difficulties arise in using this approach to study the achievement of graduates from the Modular Degree Programme. One problem is that the module averages and standard deviations needed to standardise students' marks are estimated with varying degrees of accuracy, as some modules enrol more students than others. For the module entries studied here, 12.0% of the responses were achieved in modules which had enrolled 20 or fewer students, so that the module averages and standard deviations for these modules have relatively large standard errors and would tend to make the standardised marks unreliable. A second problem is that responses are usually standardised with reference to a particular population: in a classical repeated measures design, responses on each occasion are often standardised with reference to the same population. In the context studied here, there is no single population from which all modules enrol students: each module draws students from a population defined by fields studied or by the module's pre-requisite requirements. As a result, a student's performance in the modules taken could be standardised with reference to a number of different populations.

An alternative to standardisation is to analyse the raw marks. This will lead to an exaggerated view of the variation in the performance of individual students, since the use of different scales of measurement in different modules, will tend to increase within-student variation in assessed performance. Setting aside the problem of having responses measured on many different scales, an advantage of this approach is that it involves the same assumptions as those embodied in the Modular Programme's definitions of pass/fail and grade thresholds, in the system for credit accumulation and the method for determining degree classifications. In effect, analysing raw marks means accepting the marks awarded in different modules as equivalent, in the same way as one might agree, following Cresswell (1996), to accept examinations as having comparable standards within a specific context. Working with the raw marks has the advantage of producing parameter estimates in the same units as those used within the Modular Programme and this makes it easier to see how differences in the marks achieved within a module by students belonging to different sub-groups or in different types of modules might influence degree awards and other decisions based on a student's performance.

The next sections describe models for the analysis of raw marks, based on the hierarchical structure shown in Figure 4.1, starting with a simple variance components model.

4.3 Variance components model

Using the hierarchical representation shown in Figure 4.1, the simplest three level model which can be fitted to the assessment data is a variance components model, which assumes that marks vary randomly from student to student and that a student's performance will vary randomly from term to term and module to module but incorporates no independent variables. This model can be written as:

(4.1) $$y_{ijk} = \beta_0 + v_k + u_{jk} + e_{ijk}$$

where y_{ijk} = mark achieved in the i^{th} module taken by student k in term j

β_0 is a fixed parameter and v_k, u_{jk}, e_{ijk} are the student, term and module level residuals. These are independently distributed, with $v_k \sim N(0,\sigma_v^2)$, $u_{jk} \sim N(0,\sigma_u^2)$ and $e_{ijk} \sim N(0,\sigma_e^2)$.

So that:

$\mathrm{var}(y_{ijk}) = \sigma_v^2 + \sigma_u^2 + \sigma_e^2$

$\mathrm{cov}(y_{ijk}, y_{i'jk}) = \sigma_v^2 + \sigma_u^2$ (same student and term, different module)

$\mathrm{cov}(y_{ijk}, y_{i'j'k}) = \sigma_v^2$ (same student, different terms and modules)

and

$\mathrm{cov}(y_{ijk}, y_{i'j'k'}) = 0$ (different students, any terms and modules)

For this model: σ_e^2 represents the variation in marks achieved by an individual student in modules taken in the same term, σ_u^2 the variation in mean marks achieved by an individual student in different terms and σ_v^2 the variation in mean marks between students. It is assumed that these variances are constant and that the variation in marks at all three levels is purely random.

4.4 A more complex three-level model

The variance components model can be developed to incorporate the effects of student, programme or module characteristics on the marks awarded. In this section, equation (4.1) is modified to incorporate :

- the fixed effects of term and module characteristics on mean marks, and student characteristics
- a longitudinal element or 'growth curve', varying between students, representing the student's development during the course of their degree programme
- a complex variance structure at level 1
- random effects allowing the effects of explanatory variables (in addition to term) to vary between individuals

Adding the fixed effects of student and module characteristics and of time to the variance components model, leads to equation (4.2).

(4.2) $y_{ijk} = \mathbf{X}_{ijk}\boldsymbol{\beta} + v_k + u_{jk} + e_{ijk}$

where $\boldsymbol{\beta}$ is a *(P+1)* x *1* vector of parameters $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_p$

and $\mathbf{X}_{ijk}$ is a *1* x *(P+1)* row vector $(x_{0ijk}, x_{1ijk}, \ldots x_{Pijk})$, whose elements are the values of explanatory variables for unit *(i,j,k)* and other terms are as defined earlier.

X_0 is a dummy variable whose value is constant and equal to 1 for all units: this variable introduces the intercept term, β_0. Other explanatory variables include student and module characteristics, term and powers of term. These allow the effects of time to be represented by a polynomial function of the term in which modules were taken. In some cases, explanatory variables may be transformed or centred before being included in the module, and others may represent interactions between variables.

In the next stage of model building, complex variation is introduced at level 1, allowing the variation at level 1 to become a function of the explanatory variables. This leads to equation (4.3):

(4.3) $y_{ijk} = \mathbf{X}_{ijk}\boldsymbol{\beta} + v_k + u_{jk} + \mathbf{Z}_{ijk}\mathbf{e}_{ijk}$

where $\mathbf{Z}_{ijk}$ and $\mathbf{e}_{ijk}$ are new terms, but other terms are defined as before.

$\mathbf{e}_{ijk}$ is a vector of $Q+1$ residuals, $(e_{0ijk}, e_{1ijk}, \ldots e_{Qijk})$ with mean vector $\mathbf{0}$ and $\mathbf{E}\lfloor \mathbf{e}_{ijk}\mathbf{e}_{ijk}' \rfloor = \Omega_e$.

e_{0ijk} is Normally distributed with zero mean and variance σ^2_{e0} and is related to the other residuals, with $\operatorname{cov}(e_{0ijk}, e_{qijk}) = \sigma_{e0q}$, and these other residuals $\{e_{qijk}, p = 1,..Q\}$ are defined as having zero mean and variance and as independent of each other. The elements of Ω_e are the variances and covariances between the elements of $\mathbf{e}_{ijk}$. Ω_e has non-zero elements in row 1 and column 1 and zero elements everywhere else.

$\mathbf{Z}_{ijk}$ is a row vector of the values of $Q+1$ explanatory variables for unit *(ijk)*, the first of these variables is X_0 and the others may be some or all of the variables featured in $\mathbf{X}_{ijk}$. The level 1 variation for unit (*ijk*) is now equal to $Z\Omega_e Z'$.

Explanatory variables in Z contribute to level 1 variation through covariance terms $\{\sigma_{e0q}, q = 1,\ldots Q\}$, and for example, when X_q is a dummy variable identifying units belonging to a certain category, then belonging to this category adds $2\sigma_{e0q}$ to the level 1 variation.

In the next stage of model building, coefficients of the explanatory variables $\{X_{pijk}, p = 1,2\ldots P\}$ are allowed to vary randomly at level 3 (students). This leads to equation (4.4) :

(4.4) $y_{ijk} = \mathbf{X}_{ijk}\boldsymbol{\beta} + \mathbf{W}_{ijk}\mathbf{v}_k + u_{jk} + \mathbf{Z}_{ijk}\mathbf{e}_{ijk}$

where $\mathbf{v}_k$ is a vector of student level residuals $(v_{0k}, v_{1k} \ldots v_{Sk})$

and $\mathbf{W}_{ijk}$ is a row vector whose elements record the values of a subset of explanatory variables for unit *(ijk)* and other terms are defined as before. The student level residuals are assumed to be Normally distributed with zero means and variances $\{\sigma^2_{vs}, s = 0,1...S\}$ and to be related to each other with covariances $\{\mathrm{cov}(v_{sk}, v_{tk}) = \sigma_{vst}, s,t = 0,1,...S, s \neq t\}$.

The first variable in $\mathbf{W}_{ijk}$ is the dummy variable which is constant and equal to 1 for all units. The variances and covariances, σ^2_{vs} and σ_{vst} can be used to represent two kinds of effects: when $\sigma^2_{vs} = 0$ and $\{\sigma_{vst} = 0, \text{for } t \neq 0\}$ the variable W_s produces complex variation at student level. When $\sigma^2_{vs} \neq 0$, variable W_s produces random effects at student level, so that for student k the impact of W_s on their mean marks is $\beta_{sk} = X_{s'ijk}\beta_{s'} + v_{sk}W_{sijk}$ where s' is the subscript for variable W_s in $\mathbf{X}_{ijk}$.

For equation (4.4): $\quad var(y_{ijk}) = \mathbf{W}_{ijk}\Omega_v\mathbf{W}'_{ijk} + \sigma^2_u + \mathbf{Z}_{ijk}\Omega_e\mathbf{Z}'_{ijk}$

The next section presents the results of fitting models (4.1) to (4.4) to the student record data using the multilevel modelling software, Rasbash et al (2000). This software was used to obtain maximum likelihood estimates of the parameters, using iterative generalised least squares (IGLS). Tables of parameter estimates are presented at each stage. The values of the parameter estimates obtained will be discussed in detail once the final model has been selected. In discussing the output from the analyses, the subscripts *0, p, q* and *s* used above will be replaced by the names of the relevant variables.

4.5 Results of fitting hierarchical models to student record data

Analysing the students' performances as recorded within a three level hierarchy (496 students, 4450 terms, 14,315 entries) a variance components model was fitted , leading to the parameter estimates shown in Table 4.1. In the analyses reported below, the dummy variable whose value is constant and equal to 1 for all units was labelled 'CONS'. Parameters with the subscript 0 in the previous section and associated with this variable are labelled 'CONS' in the tables of results.

Table 4.1 Parameter Estimates for Variance Components Model (4.1)

Parameter	Estimate	Standard error
Fixed:		
Constant	57.55	0.2658
Random:		
σ_v^2 between students	31.91	2.227
σ_u^2 between terms	6.047	0.6931
σ_e^2 between module entries	70.04	0.9914
-2*log(lh) = 103732		

This shows that the greatest variation occurs at level 1, between module entries, followed by variation between students, at level 3. To develop the model, variables measured at student, term and module level were introduced as fixed effects. These explanatory variables are listed in Table 4.2.

Table 4.2 Explanatory variables used in analyses

CONS	= 1 for all units
Variables measured at student or programme level:	
MATURE	1= aged 21 or over at enrolment 0 = other
MALE	1 = male 0 = female
PROFMAN	1= parent with professional/managerial occupation 0 = other
INTERNAT	1= domicile outside the UK 0 =other
OFFER	= number of A-level points for typical offer for student's chosen degree award
COFFER	= OFFER, centred about the mean
Variables measured at term level:	
TERM	1-9
CTERM	= TERM – 4.5
FIVEPLUS	1 = student enrolled on 5 or more modules in current term 0 = other
Variables measured at module level	
Modules were divided into 14 schools/subject groups and dummy variables used to identify modules in categories other than social sciences. Social science modules defined a reference category against which others are compared.	
VAMP	1 = visual arts, music and publishing 0 = other
BMS	1=Biology and molecular sciences 0 = other
BUSINESS	1= Business 0 = other
CMS	1 = computing and mathematical sciences 0 = other
CON+ES	1 = construction and earth science 0=other
EDUCATION	1= education 0=other
ENGINEER	1 = engineering 0=other
H+RM	1 = hotel and restaurant management 0=other
HUMAN	1 = humanities 0=other
LANG	1 = languages 0=other
PLANNING	1 = Planning 0=other
LAW	1 = Law 0=other
UNATTACHED	1= modules not attached to a school/department 0=other
ALLCSW	1 = 100% coursework assessment 0=other
SOMECSW	1 = mixed coursework and examination assessment 0=other
CRTN	Square root of number of students enrolled, centred
ST2BASIC	1 = basic module taken in stage 2 0=other
DOUBLE	1= double module 0=other
PROJECT	1 = project module 0=other

4.5.1 A Main Effects Model

The explanatory variables listed in Table 4.2 were used to fit a fixed, main effects model corresponding to equation (4.2). Most of these variables are dummy variables taking the values 1 or 0 and are used to introduce to the model the effects on performance of units of belonging to certain categories. A small number of variables, measures of class size, entry qualifications and term have been centred so that the intercept term refers to performance when these variables have their mean values, rather than when these measure are equal to zero. Centring makes the intercept terms more meaningful when zero values are unrealistic.

The variables measured at student level record each student's age, sex, domicile, (parent's) social class and the A-level points required for their choice of degree. The last of these is the number of points corresponding to a typical offer for the student's chosen field(s). This measures the level of achievement 'expected' of a student rather than the actual level achieved (see earlier discussion in section 3.9.2). This variable was centred (COFFER) before being introduced to the model.

Two independent variables were available at level 2; term and workload, represented by a dummy variable identifying terms in which the student had entered five or more modules, as this represents an unusually high workload. The variable 'term' (running from 1 to 9) was centred around term 5 (CTERM) and the effects of term on mean marks included in the module as a fourth degree polynomial, with higher order terms found not to make a significant contribution to the model.

Variables measured at module level identify which of 14 schools or departments are responsible for teaching and assessment, the method of assessment used, the number of students enrolled, the type (project or other) and level of the module and the number of module credits it awards (single or double). The explanatory variables include only one indicator of 'level' (ST2BASIC), identifying basic modules taken in stage 2. The levels of the modules taken in a student's programme are largely determined by the year in which they are taken. Only basic modules are taken in stage 1 and in stage 2 a student's programme consists mainly of advanced modules. If 'term' is modelled explicitly, then all the remaining variation in level can be identified by a variable which identifies basic modules taken in stage 2. Module enrolment figures were positively skewed, with one or two exceptionally large modules; a square root transformation produced a linear relationship between module marks and (root) class size and the transformed values were centred about the mean of 7.837. Centring class size means that the constant term, β_{cons}, refers to marks achieved in modules with a 'typical' enrolment of 61.

The intercept term, β_{cons}, represents the mean mark achieved in module entries for which all the explanatory variables, other than CONS, are zero. Variances σ^2_{vcons}, σ^2_u and σ^2_{econs} are the student, term and module entry level variances for units for which all the explanatory variables, other than CONS, are zero. These parameters refer to 'standard' units: female students who entered the course at the traditional age, whose parents are not in professional or managerial occupations, domiciled in the UK and graduate in subjects for which the specified entry qualifications are 15.7 A-level points. Standard module entries are in single, social science modules, which are not final year project or dissertation modules, using 100% examination assessment, have 61.4 students enrolled, and taken by 'standard' students with a workload of less than five modules per term. Standard modules are taken the hypothetical 4.5th term of a student's programme. Standard students and module entries define a reference category against which the marks achieved by other types of students or in other module categories will be compared.

Table 4.3 shows the parameter estimates obtained when the variables listed above were used to fit model (4.2) to the student record data. The introduction of these variables as fixed effects reduces the total deviance by 830 with 29 degrees of freedom ($P < 10^{-15}$) and leads to some reduction in the estimated variation at each level. Estimated effects of student characteristics show that mature students tend to perform better than those who start their degree at the traditional age, while on average, international students achieve lower marks than home students. The estimated fixed effects of module characteristics on mean achievement show that mean marks are lower for projects than for other 100% coursework advanced modules taken at the same time. The use of coursework assessment is associated with higher mean marks, other things being equal, and the increase in the mean is greater when coursework is the only method of assessment used in a module. There were significant changes in mean marks over time and these are discussed in more detail below.

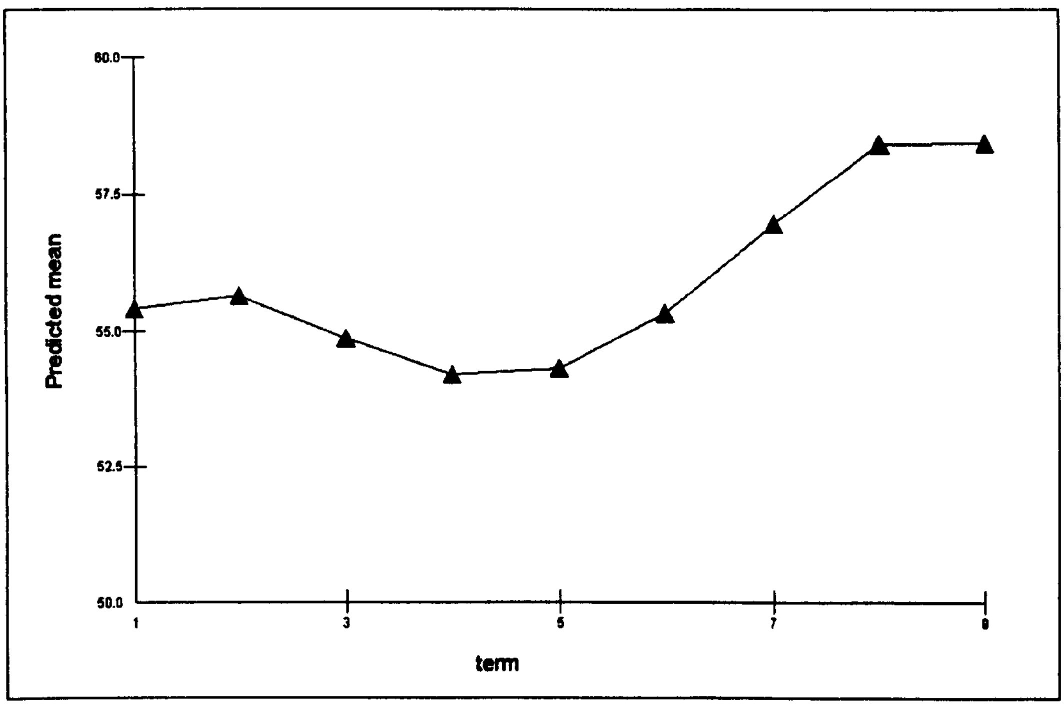
Table 4.3 Parameter estimates for fixed, main effects model (4.2)

Parameter	Estimate	Standard error
Fixed:		
Constant	54.32	0.6387
Project	-2.074	0.3463
Male	-0.7148	0.529
international	-2.325	0.7371
mature	2.925	0.6162
coffer	-0.09342	0.06188
Professional/managerial	0.6208	0.5368
Workload: 5+	-0.5137	0.3938
cterm	0.5767	0.08093
$cterm^2$	0.481	0.05269
$cterm^3$	-0.01239	0.006132
$cterm^4$	-0.01984	0.003083
Assessment method:		
100%csw	2.302	0.309
Mixed	1.107	0.2861
Subjects:		
vamp	0.9498	0.4799
bms	-2.242	0.4271
business	1.515	0.3859
cms	-0.8694	0.4155
con+es	-0.7929	0.5537
educ	1.422	0.6923
engineer	-0.9057	1.214
h+rm	2.682	0.897
human	-0.1511	0.3885
lang	-3.53	1.005
planning	1.579	0.5509
law	-5.748	0.5184
unattach	0.2113	1.005
Double	0.4937	0.2195
Root class size (centred)	0.01553	0.0296
Stage 2 basic	-0.1204	0.4818
Random:		
σ_v^2 between students	30.05	2.088
σ_u^2 between terms	4.041	0.6293
σ_e^2 between module entries	67.42	0.9536
-2*log(lh)	102905	

Although the estimated effects of some variables included in the model did not produce statistically significant reductions in the likelihood ratio statistic, these variables have been retained as they may be useful when other kinds of effects are introduced into the model, for example interactions or complex variance structures.

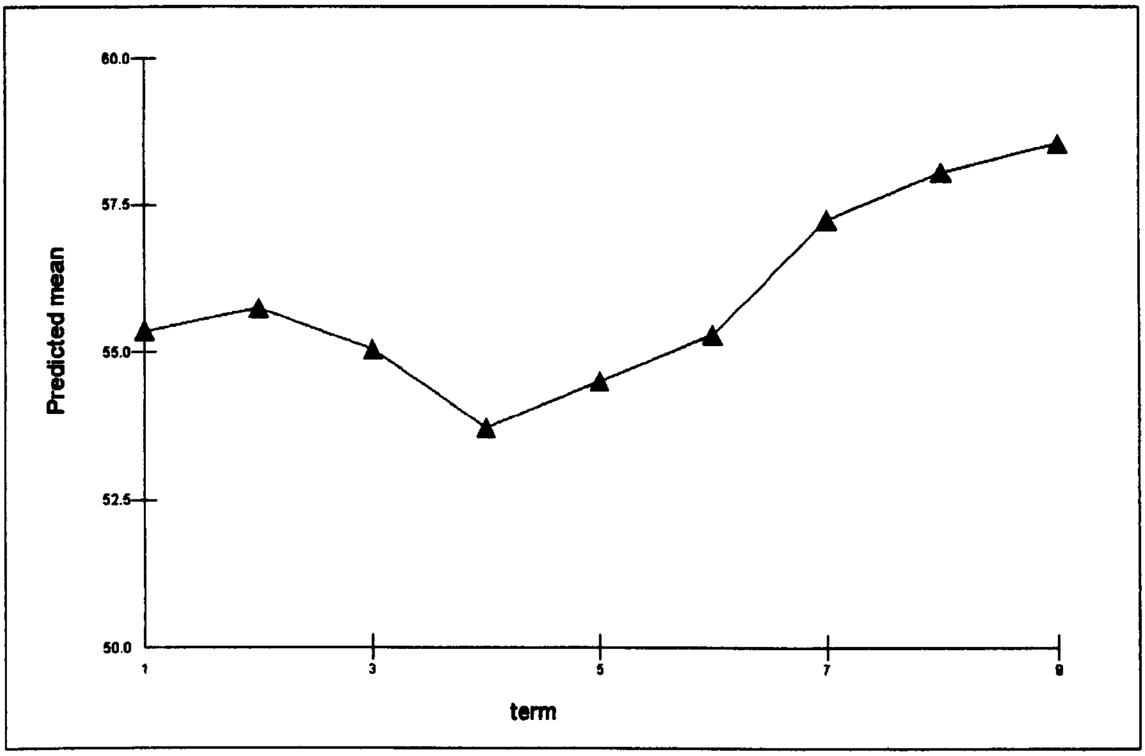
In Table 4.3, mean progress over time is represented by a polynomial curve of degree 4. For module entries with other variables equal to zero, the predicted mean mark in term t is $\beta_0 + \beta_{t_1}(t-4.5) + \beta_{t_2}(t-4.5)^2 + \beta_{t_3}(t-4.5)^3 + \beta_{t_4}(t-4.5)^4$, where β_{t_i} is the coefficient of the variable $CTERM^i$ .Higher order terms do not make a statistically significant contribution to the model: for example, the estimated regression coefficient of $CTERM^5$ is 0.003 with standard error 0.002. The fourth degree polynomial has a number of turning points, as shown in Figure 4.2

Figure 4.2 Mean marks by term, model (4.2)



As the shape of this curve is relatively complex and introduces a number of parameters to the model, alternative forms for representing mean progress were investigated by fitting a model using dummy variables to incorporate an individual mean for each term. This model avoids any assumptions about the shape of the mean progress curve. The polynomial terms in (CTERM) were replaced by dummy variables for each term. For this model, the estimated mean marks for each term for students for whom all other independent variables are zero are shown in Figure 4.3. The shape of this curve suggests that students' progress follows different patterns in stages 1 and 2 (year 1 / years 2 and 3).

Figure 4.3 Fitted progress curve using dummy variables to represent terms

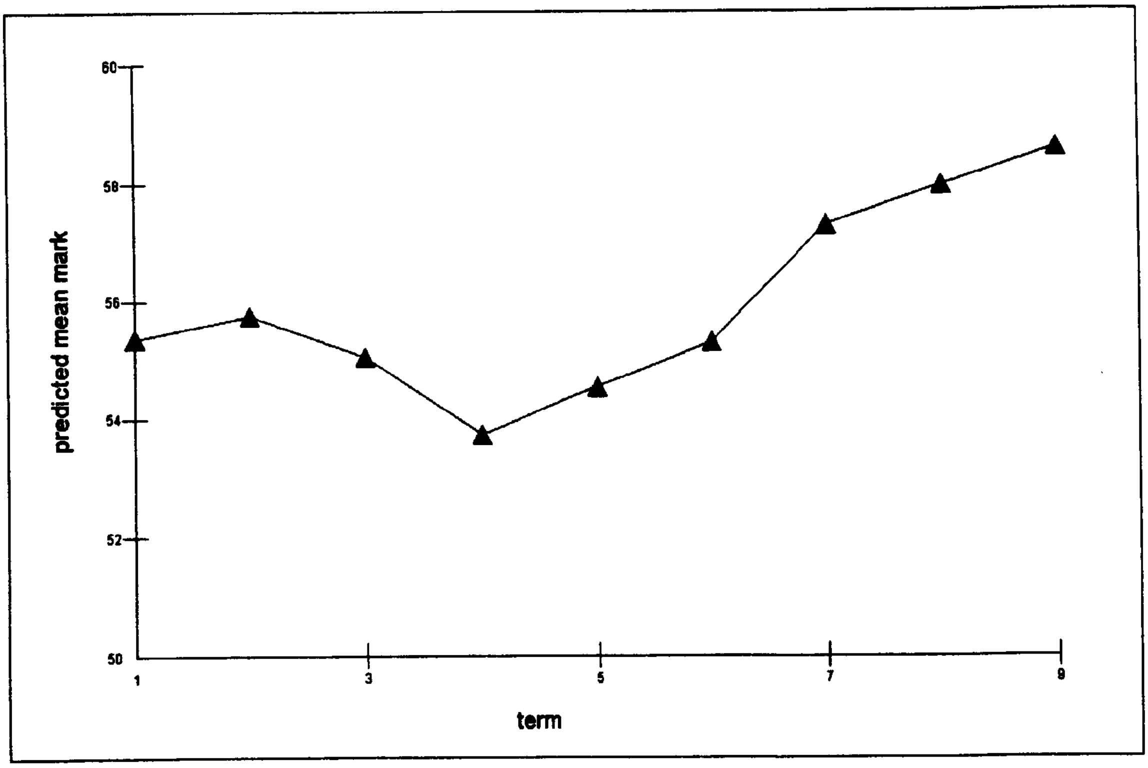


Within stage 2, mean progress from term to term can be modelled as a polynomial function of 'term'. The graph suggests a potential discontinuity between years 2 and 3: if this is the case, then it would be useful to be able to estimate this directly by including a shift from year 2 to year 3 within the model. Replacing the powers of CTERM by new variables leads to predicted mean marks shown in Figure 4.4. The variables used to achieve this model of mean progress are:

YEAR1	1= modules taken in first year	0 = other
YEAR3	1= modules taken in third year	0 = other
TERM1	1= modules taken in term 1 of the first year	0 = other
TERM3	1= module taken in term 3 of the first year	0 = other
ST2LINEAR	= (TERM-6.5)*(1-YEAR1)	
ST2QUAD	= $(\text{ST2LINEAR})^2$	

These variables are used to fit a model of progress describing mean achievement as varying from term to term in the first year, with a change between mean achievement between the first two terms and the third term of the first year and a quadratic model for progress in stage 2, with the addition of a 'step' between years 2 and 3. Figure 4.4 shows the predicted mean progress term by term for this model. The graph shows the same pattern of progress as in Figure 4.3, but is achieved with fewer parameters (six rather than eight), with the potential for non-significant terms to be dropped at a later stage.

Figure 4.4 Revised model for mean progress



The parameter estimates obtained for the main effects model using these variables to model progress are shown in Table 4.4. The results in Table 4.4 indicate that a simpler model of students' progress would be adequate, since the coefficient of the quadratic term does not significantly reduce the likelihood ratio statistic for the model, but fitting a more complex model than is strictly necessary is worthwhile at this stage to ensure that, at the next stage, the full range of interactions between 'term' and other independent variables can be explored.

Table 4.4 Parameter estimates for main effects model with revised model for progress

Parameter	Estimate	Standard error
Fixed:		
Constant	55.66	0.6714
Project	-2.01	0.3482
Male	-0.7149	0.5287
international	-2.325	0.7368
mature	2.924	0.6159
coffer	-0.09292	0.06185
Professional/managerial	0.622	0.5365
Workload: 5+	-0.637	0.394
Year1	0.0901	0.338
Term1	-0.384	0.3257
Term3	-0.7092	0.3162
Year3	1.284	0.393
Stage 2 linear	0.7184	0.1147
(Stage 2 linear)^2	-0.02329	0.03754
Assessment method:		
100%csw	2.291	0.3103
Mixed	1.115	0.2871
Subjects:		
vamp	0.9309	0.4798
bms	-2.236	0.4271
business	1.529	0.3859
cms	-0.8422	0.4154
con+es	-0.8043	0.5536
educ	1.428	0.6922
engineer	-0.8698	1.214
h+rm	2.657	0.8971
human	-0.1313	0.3885
lang	-3.647	1.006
planning	1.578	0.5507
law	-5.719	0.5187
unattach	0.1081	1.005
Double	0.4897	0.2202
Root class size (centred)	0.01036	0.02976
Stage 2 basic	-0.001172	0.4836
Random:		
σ_v^2 between students	30.03	2.086
σ_u^2 between terms	3.986	0.6281
σ_e^2 between entries	67.41	0.9535
-2*log(lh)	102895	

4.5.2 A model including interactions between independent variables

At the next stage, interaction terms were added to the model: two-way interactions were introduced and tested individually as additions to the main effects model estimated in Table 4.4. Interactions with time were assessed by fitting the interactions with 'YEAR1', 'YEAR3', 'ST2LINEAR' and 'ST2QUAD'. The interactions tested were :

gender by age, parents' social class, 'offer', assessment method, project, class size, subject, time

age by parents' social class, 'offer', assessment method, project, class size, subject, time

parents' social class by time

class size by assessment method, subject, time

assessment method by time and subject.

Although 14 subject groups and three assessment strategies were identified (100% coursework, 100% examination or a mixture), in some subject groups module entries were concentrated within only one or two assessment categories hence 15 dummy variables (rather than 26) were required to introduce this interaction.

Interactions which made a statistically significant contribution to the model, after allowing for main effects, were added collectively. Interactions selected at this stage were: gender x class size , age x assessment method, class size x assessment method , gender x time, age x time, parents' social class x time, class size x time, assessment method by time , class size by subject and assessment method x subject. The contributions of each dummy variable representing these interactions were tested after allowing for main effects and all other interaction effects. Those which did not contribute significantly to the model were removed, leading to a reduced interaction model. Dummy variables which had been removed were re-tested as additions to the reduced model and some reintroduced at this stage. The interactions included in the final interaction model are: parents' social class x year3, gender x some csw, class size x yr1, gender x year1, mature x year3, mature x st2linear, mature x somecsw, all csw x bms, business, cms, eng, h+rm, human, langs, law , somecsw x cms, law, class size x bms, bus, cms, con +es. When reintroduced collectively, some estimated model parameters are small relative to their standard errors; the contribution of these terms are reviewed after incorporating a complex random structure into the model. As the quadratic term describing progress in stage 2 did not feature in the selected interactions, it was removed from the main effects at this stage.

Parameter estimates for the final interaction model are shown in Table 4.5. This model adds 23 new parameters to the main effects model and reduces the total deviance by 328.0. The estimated interaction effects imply that the progress made by students in different categories follows different patterns, and that the impact of assessment methods also differs for students in different categories. The values of the parameter estimates will be discussed in greater detail once the model is complete.

Table 4.5 Parameter estimates for model including main effects and interactions

Parameter	Estimate	Standard error
Fixed:		
Constant	55.25	0.6859
Project	-2.643	0.3609
Male	-0.8233	0.5628
International	-2.319	0.7349
Mature	3.566	0.6765
Coffer	-0.1155	0.06176
Professional/managerial	0.3745	0.5461
Workload: 5+	-0.6966	0.3893
Year1	-0.242	0.3483
Term1	-0.3789	0.3255
Term3	-0.7377	0.3131
Year3	0.8262	0.4659
Stage 2 linear	0.82	0.1232
Assessment method:		
100%csw	2.02	0.3932
Mixed	2.181	0.3611
Subjects:		
vamp	0.8677	0.4885
bms	-2.902	0.4491
business	0.8949	0.4059
cms	-1.481	1.348
con+es	-1.639	0.6331
educ	1.52	0.7038
engineer	-3.07	1.354
h+rm	0.7992	1.142
human	-0.6817	0.4502
lang	-12.48	1.534
planning	1.57	0.5478
law	-3.592	0.8002
unattach	0.09252	1.004
Double	0.8119	0.2271
Root class size (centred)	-0.2299	0.06647
Stage 2 basic	0.4136	0.4888

Interactions:		
Profman x year3	0.8742	0.3287
Size x year1	0.1265	0.05459
Allcsw x bms	3.319	0.7469
Allcsw x bus	0.8236	0.472
Allcsw x cms	-2.131	1.423
Allcsw x engineer	7.448	2.204
Allcsw x h+rm	3.441	1.375
Allcsw x human	1.149	0.5157
Allcsw x lang	13.94	1.919
Allcsw x law	1.567	0.8786
Mature x somecsw	-1.186	0.3578
Size x bms	0.2596	0.1049
Size x bus	0.3236	0.07696
Size x cms	0.3046	0.1174
Size x c+es	-0.3334	0.1989
Size x h+rm	0.7347	0.2971
Size x law	-0.4234	0.1394
male x year1	1.319	0.3223
mature x year3	-0.2894	0.5665
mature x st2linear	-0.4273	0.1917
somecsw x cms	1.511	1.35
some csw x law	-1.881	0.7754
male x somecsw	-0.6993	0.3098
Random:		
σ_v^2 between students	29.93	2.075
σ_u^2 between terms	3.78	0.6117
σ_e^2 between module entries	65.94	0.9326
-2*log(lh)	102567	

4.5.3 A model including a complex variance structure at level 1

In the third stage of model building, a complex variance structure was introduced at level 1, so that the model fitted corresponds to equation (4.3). This development allows the level 1 variation to become a function of independent variables measured at any level in the model. Variables introduced at this stage were: assessment method, subject, workload, time (YEAR1, YEAR3), gender, age, domicile, social class and the interaction between gender and method of assessment. Estimates of the elements of Ω_e which, after allowing for other variables, did not significantly reduce the likelihood ratio statistic were removed and terms introducing different levels of variation by subject were streamlined by combining subjects with similar level 1 variances. The final subset of variables defining level 1 variation were : year1, year3, gender, age, assessment method, subject (bms, business, cms, con+es, engineering, languages, law) , workload and domicile. The parameter estimates for this model are shown in Table 4.6.

With complex variation at level 1, the estimate of σ_u^2, the variation between terms falls from 3.78 (se=0.612) in Table 4.5 to 2.79 (se = 0.517) in Table 4.6. The estimates of random parameters at level 1 show that the use of coursework assessment leads to a reduction in the variation between module entries. There are subject differences in the variation between module entries, with sciences (biology and molecular sciences, computing and mathematical sciences, engineering) having the highest levels of variation at level 1. Language modules also have very high level 1 variation but since the sample selection excludes language students, these results are based entirely on module entries made by students studying for non-language degrees and are therefore drawn from an unusual population. Students who are male, international students or who entered the degree course at the traditional age tend to have greater variation at level 1. Level 1 variation appears to decrease from one year to the next: the greatest variation occurring in the first year and the lowest in the final year.

Table 4.6 Parameter estimates for model (4.3) including complex variation at level 1

Parameter	Estimate	Standard error
Fixed:		
Constant	55.2	0.678
Project	-2.829	0.3048
Male	-0.7864	0.5463
International	-2.379	0.7169
Mature	3.544	0.6591
Coffer	-0.1201	0.0604
Professional/managerial	0.3738	0.5336
Workload: 5+	-0.5951	0.3806
Year1	-0.2793	0.3287
Term1	-0.5052	0.3306
Term3	-0.7982	0.3175
Year3	0.7193	0.4207
Stage 2 linear	0.8257	0.1111
Assessment method:		
100%csw	1.98	0.3987
Mixed	2.103	0.3776
Subjects:		
vamp	0.6741	0.4255
bms	-2.415	0.4363
business	1.052	0.3826
cms	-1.078	2.044
con+es	-1.976	0.6546
educ	1.535	0.5924
engineer	-2.473	1.699
h+rm	0.821	0.9433
human	-0.7688	0.393
lang	-12.71	2.515
planning	1.602	0.4778
law	-2.983	0.892
unattach	0.4685	0.9189
Double	0.8798	0.1992
Root class size (centred)	-0.225	0.05658
Stage 2 basic	0.03664	0.4989
Interactions:		
Profman x year3	0.9637	0.2966
Size x year1	0.153	0.05372
Allcsw x bms	3.376	0.7705

Allcsw x bus	0.8791	0.4268
Allcsw x cms	-1.955	2.147
Allcsw x engineer	7.569	2.817
Allcsw x h+rm	3.296	1.145
Allcsw x human	0.7752	0.4323
Allcsw x lang	14.13	3.183
Allcsw x law	1.51	0.9553
Mature x somecsw	-1.166	0.3268
Size x bms	0.2661	0.1046
Size x bus	0.3133	0.06894
Size x cms	0.246	0.1469
Size x c+es	-0.3902	0.2088
Size x h+rm	0.6653	0.2377
Size x law	-0.4847	0.135
male x year1	1.043	0.3194
mature x year3	0.0246	0.5186
mature x st2linear	-0.489	0.1742
somecsw x cms	1.546	2.068
some csw x law	-1.857	0.8686
male x somecsw	-0.5281	0.2966
Random:		
σ_v^2 between students	28.53	1.967
σ_u^2 between terms	2.792	0.5169
Level 1 variance σ^2_{econs}	76.84	4.149
Covariance between level 1 residuals associated with cons and:		
100% coursework	-16.98	2.031
mixed assessment	-18.97	1.998
bms	16.55	1.499
bus	6.048	0.9322
cms	46.22	3.19
con+es	15.41	2.028
engineer	32.1	9.762
lang	73.54	15.5
law	8.981	1.542
workload 5+	6.089	1.864
male	4.639	0.745
international	3.534	1.062
mature	-2.088	0.7577
year1	5.028	0.9542
year3	-3.562	0.8033
-2*log(lh)	101536	

4.5.4 A model including effects defined as random at level 3

Finally, equation (4.4) was implemented by allowing some regression coefficients to vary at level 3. This introduced some individualised effects and a complex variance structure at level 3. The effects of assessment method, class size and time were allowed to vary from student to student and the individual level residuals associated with these variables were allowed to covary. Other terms were introduced at level 3 to allow between-student variation to differ for male and female students and between students of different ages.

With the introduction of individual coefficients for YEAR1 and YEAR3, variation at level 2, between terms, was completely explained. Estimated variation between students in the linear element of their progress during years 2 and 3 is not significant ($\hat{\sigma}^2_{v.st2linear} = 0.103, se = 0.1095 : \chi^2_1 = 1.174, P > 0.05$).

Estimates of the parameters describing variation at level 3 that did not contribute significantly to the model were removed, then reintroduced and re-tested individually. The final selection of parameters describing the variation at level 3 is shown below, arranged as the lower triangular elements of the matrix $\Omega_v = E[\mathbf{vv}']$, where $\mathbf{v}$ is the vector of student level residuals. The corresponding parameter estimates presented in Table 4. 7.

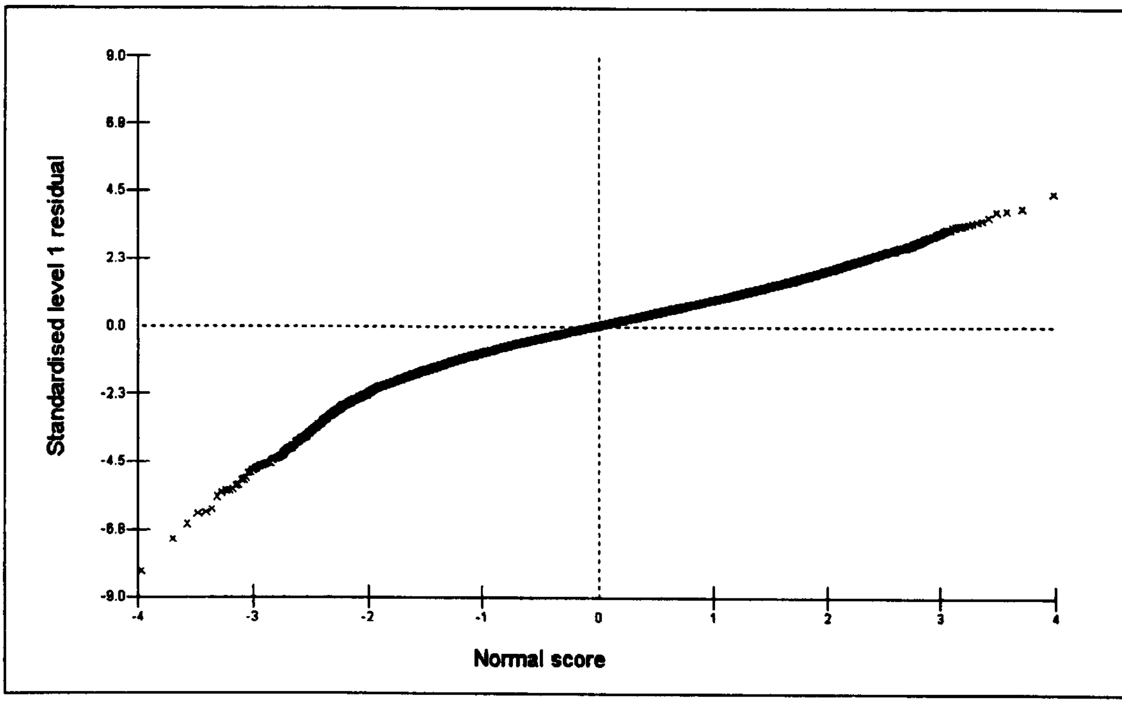
$$\Omega_v = \begin{bmatrix} \sigma^2_{vcons} & & & & & \\ \sigma_{vcons/allcsw} & \sigma^2_{vallcsw} & & & & \\ \sigma_{vcons/somecsw} & \sigma_{vsomecsw/allcsw} & \sigma^2_{vsomecsw} & & & \\ 0 & \sigma_{vyear1/allcsw} & \sigma_{vyear1/somecsw} & \sigma^2_{vyear1} & & \\ 0 & 0 & 0 & 0 & \sigma^2_{vyear3} & \\ \sigma_{vcons/mature} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Table 4.7 Parameter estimates for equation (4.4) including random effects and complex variation at level 3

Parameter	Estimate	Standard error
Fixed:		
Constant	54.99	0.6797
Project	-2.837	0.3002
Male	-0.8322	0.5359
International	-2.112	0.692
Mature	3.609	0.6755
Coffer	-0.1361	0.05877
Professional/managerial	0.1361	0.051
Workload: 5+	-0.228	0.371
Year1	-0.2337	0.3456
Term1	-0.5094	0.3068
Term3	-0.7844	0.2926
Year3	0.8144	0.456
Stage 2 linear	0.7979	0.1032
Assessment method:		
100%csw	2.291	0.4454
Mixed	2.328	0.3987
Subjects:		
vamp	0.6883	0.4233
bms	-2.402	0.4396
business	1.25	0.3892
cms	-0.7396	2.037
con+es	-1.979	0.6474
educ	1.37	0.6011
engineer	-2.26	1.674
h+rm	0.6197	0.9607
human	-0.7199	0.4029
lang	-12.79	2.465
planning	1.526	0.4701
law	-2.713	0.9402
unattach	0.03246	0.9149
Double	0.8195	0.2004
Root class size (centred)	-0.2089	0.05708
Stage 2 basic	0.1093	0.4963
Interactions:		
Profman x year3	0.967	0.3899
Size x year1	0.1473	0.05499
Allcsw x bms	3.297	0.7931
Allcsw x bus	0.6648	0.4683
Allcsw x cms	-2.322	2.146
Allcsw x engineer	8.245	2.838
Allcsw x h+rm	3.166	1.22
Allcsw x human	0.4157	0.4723
Allcsw x lang	14.16	3.115
Allcsw x law	1.451	1.068
Mature x somecsw	-1.218	0.3556
Size x bms	0.2856	0.1082
Size x bus	0.2993	0.06939
Size x cms	0.2222	0.146
Size x c+es	-0.4216	0.215
Size x h+rm	.7226	0.2355
Size x law	-0.5331	0.1343
male x year1	1.099	0.3991

mature x year3 mature x st2linear somecsw x cms some csw x law male x somecsw	-0.07978 -0.4748 1.322 -1.871 -0.5172	0.6108 0.1714 2.054 0.8899 0.3244
Random: Student level:		
σ^2_{vcons}	39.69	5.13
$\sigma_{vcons/allcsw}$	-15.5	3.8
$\sigma^2_{vallcsw}$	14.3	3.56
$\sigma_{vcons/somecsw}$	-8.449	3.266
$\sigma_{vallcsw/somecsw}$	7.819	2.837
$\sigma^2_{vsomecsw}$	5.145	2.647
$\sigma_{vcons/mature}$	6.253	2.606
$\sigma_{vcyear1/allcsw}$	-3.712	0.9778
$\sigma_{vyear1/somecsw}$	-2.901	0.9201
σ^2_{vyear1}	7.081	1.168
σ^2_{vyear3}	7.683	1.074
Between terms σ^2_u	0	0
Level 1 variance σ^2_{econs}	70.16	4.135
Covariance between level 1 residuals associated with cons and:		
100% coursework	-13.39	2.032
mixed assessment	-15.13	2.002
bms	16.28	1.461
bus	5.685	0.8942
cms	45.82	3.137
con+es	14.92	1.966
engineer	29.87	9.306
lang	68.13	14.55
law	8.84	1.492
workload 5+	5.608	1.784
male	4.389	0.7159
international	2.917	1.006
mature	-1.93	0.7274
year1	3.681	0.9265
year3	-4.813	0.776
-2*log(lh)	101244	

Figure 4.5 Standardised Level 1 Residuals by Equivalent Normal Scores for Analysis shown in Table 4.7



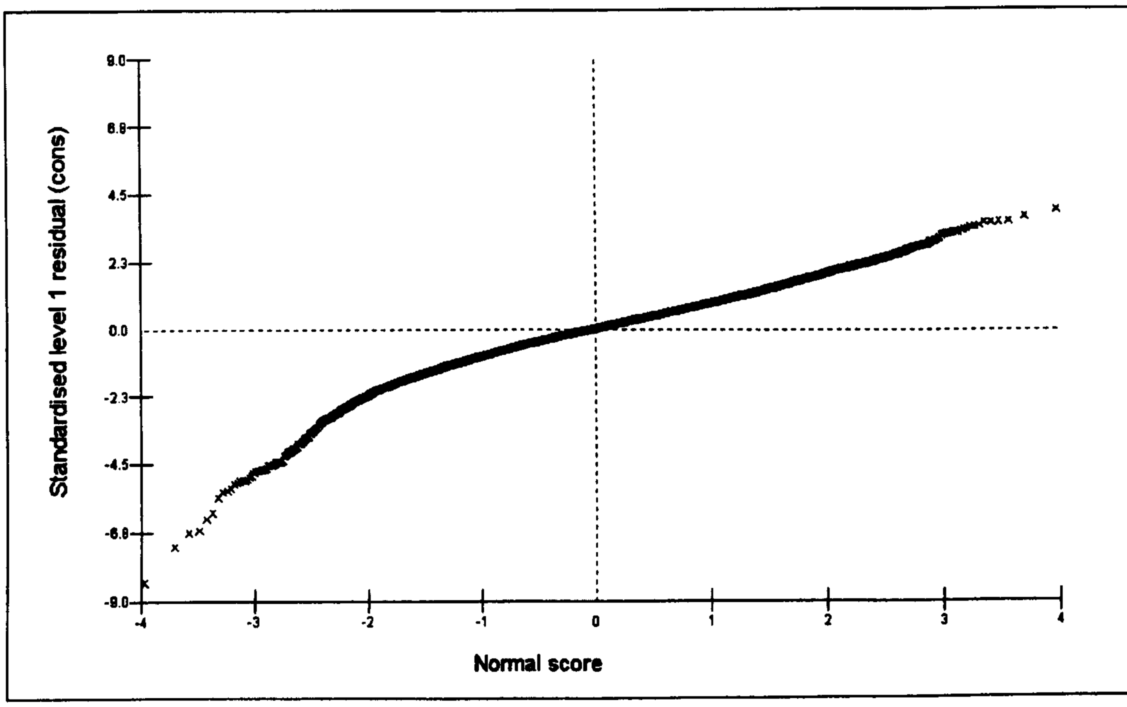
Having fitted the fixed and random effects included in equation (4.4), residual plots were used to verify the model assumptions. Figure 4.5 shows that the estimated level 1 residuals do not appear to be Normally distributed. One explanation may be that additional variables are needed to explain the variation at level 1 and so a wider range of variables was considered as potentially influencing level 1 variation. This lead to the addition of some new parameters, with parameter estimates as shown in Table 4.8, however Figure 4.6 shows that only minor improvements in the level 1 residual plot are produced.

Table 4.8 Parameter estimates after adding variables to complex variance structure at level 1

Parameter	Estimate	Standard error
Fixed:		
Constant	55.15	0.6774
Project	-2.805	0.2781
Male	-0.8412	0.5371
International	-2.105	0.6836
Mature	3.469	0.6752
Coffer	-0.1396	0.05821
Professional/managerial	0.1233	0.5046
Workload: 5+	-0.2248	0.3673
Year1	-0.2899	0.3461
Term1	-0.3981	0.3022
Term3	-0.7148	0.2921
Year3	0.7975	0.4612
Stage 2 linear	0.7781	0.1027
Assessment method:		
100%csw	2.303	0.4539
Mixed	2.181	0.4021
Subjects:		
vamp	0.7513	0.4244
bms	-2.41	0.4434
business	1.168	0.3953
cms	-0.8798	1.782
con+es	-2.29	0.6453
educ	1.169	0.6269
engineer	-2.521	1.655
h+rm	0.3597	0.9487
human	-0.7596	0.3948
lang	-12.93	2.529
planning	1.358	0.468
law	-2.834	0.9224
unattach	-0.1231	0.964
Double	0.7772	0.1982
Root class size (centred)	-0.2011	0.0563
Stage 2 basic	0.1481	0.5808
Interactions:		
Profman x year3	1.037	0.3978
Size x year1	0.1352	0.0536
Allcsw x bms	3.218	0.7499
Allcsw x bus	0.5429	0.4709
Allcsw x cms	-2.08	1.871
Allcsw x engineer	8.572	2.864
Allcsw x h+rm	3.203	1.251
Allcsw x human	0.3611	0.4827
Allcsw x lang	14.18	3.199
Allcsw x law	1.255	1.047
Mature x somecsw	-0.9965	0.3689
Size x bms	0.2731	0.1086
Size x bus	0.314	0.06845
Size x cms	0.07843	0.1394
Size x c+es	-0.5029	0.2146
Size x h+rm	0.7112	0.2354

Size x law	-0.5617	0.1307
male x year1	1.111	0.3958
mature x year3	0.06989	0.6162
mature x st2linear	-0.4882	0.1702
somecsw x cms	1.529	1.813
some csw x law	-1.778	0.8777
male x somecsw	-0.5241	0.3361
Random:		
Student level: σ^2_{vcons}	39.81	5.083
$\sigma_{vcons/allcsw}$	-15.92	3.793
$\sigma^2_{vallcsw}$	15.96	3.597
$\sigma_{vcons/somecsw}$	-9.023	3.249
$\sigma_{vallcsw/somecsw}$	7.977	2.826
$\sigma^2_{vsomecsw}$	5.805	2.638
$\sigma_{vcons/mature}$	6.069	2.559
$\sigma_{vcyear1/allcsw}$	-3.821	0.9813
$\sigma_{vyear1/somecsw}$	-3.177	0.9109
σ^2_{vyear1}	7.092	1.15
σ^2_{vyear3}	8.443	1.121
Between terms σ^2_u	0	0
Level 1 variance		
σ^2 [illegible]	66.39	4.012
Covariance between level 1 residuals associated with cons and:		
project	-6.869	0.9532
100% coursework	-13.4	2.04
mixed assessment	16.93	1.995
bms	19.99	1.671
bus	12.09	1.195
cms	23.86	3.913
con+es	14.51	1.943
engineer	28.13	9.0
lang	66	15.41
law	4.956	2.143
workload 5+	4.397	1.688
male	3.553	0.6843
mature	-1.671	0.6971
class size	-0.5516	0.1663
year1	7.66	0.9146
stage2 basic	15.83	3.891
size xyear1	-0.9437	0.2283
bms x allcsw	-15.21	3.152
bus x allcsw	-11.91	1.755
cms x somecsw	36.77	5.945
law x somecsw	9.674	2.92
-2*log(lh)	101099	

Figure 4.6 Standardised Level 1 Residuals by Equivalent Normal Scores after adding variables to complex variance structure at level 1



The non-linearity in Figure 4.6 is produced by module entries with large negative residuals, indicating recorded marks well below the expected level of performance. Within the Modular Degree Programme, once a module has started, it cannot be deleted from a student's programme, and will appear in a students record even if they have completed no assessed work. In section 3.5 it was explained that module entries for which marks of less than 5 were recorded were excluded from the sample as being unlikely to represent the outcome of a genuine attempt to complete a module. This procedure may have been too conservative, allowing some 'abandoned attempts' at passing a module to be included in the sample. Including these modules in the analyses would be expected to generate large negative residuals at level 1 and to influence parameter estimates, for example, by inflating estimated level 1 variance parameters. Identification of these entries is desirable but a cautious approach is needed to distinguish between the marks recorded for 'abandoned' module entries and other low marks achieved by students who completed the assessed work, intending that their result would contribute to their record.

Within the regulations that were applied to the cohort studied here, failed module entries with marks of below 20 are treated differently to other 'fails' since in circumstances where the opportunity to resit is potentially available, this would not normally be awarded to students whose mark is below 20, who are regarded as not having made a serious attempt at the module. Within the sample, 78 module entries had recorded marks below 20; Figure 3.6 shows that these represent the lower extreme of the mark distribution. The sample included 63 failed module entries with marks more than 3 standard deviations below the student's individual mean (using the standard deviation calculated from the individual student's record). Using these two criteria, 112 module entries were identified as potentially having been 'abandoned': 49 with marks below 20, 34 with marks indicating performance, in statistical terms, below that expected of the student and 29 identified by both criteria. These entries continued to be included in the model, but a new parameter was introduced to the fixed part of the model, by adding to the independent variables a dummy variable (labelled 'TAKEOUT') to indicate module entries fitted separately.

Figure 4.7 shows that fitting a separate term for these module entries, leads to an improvement in the level 1 residual plots and Figure 4.8 shows that the estimated level 3 residuals also have an approximately normal distribution. As variation at level 2 is completely explained, there are no level 2 residuals to examine. Parameter estimates are shown in Table 4.9.

Figure 4.7 Standardised level 1 residuals by equivalent Normal scores after identifying 112 'abandoned' module entries

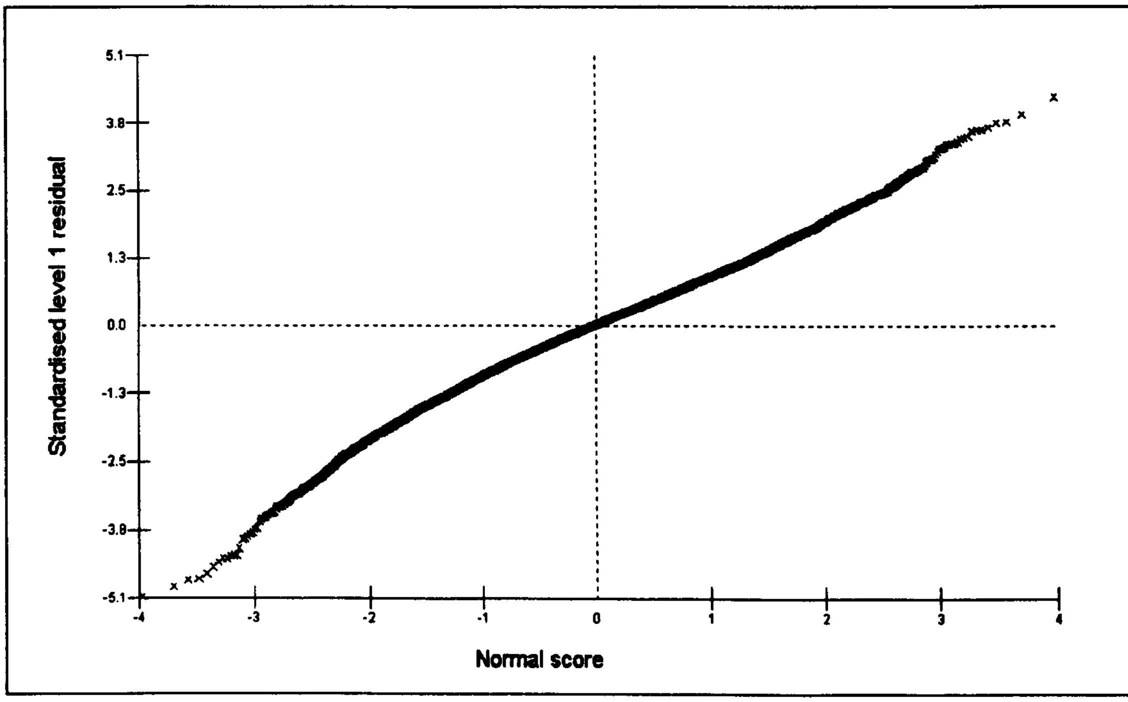


Figure 4.8 Standardised level 3 residuals by equivalent Normal scores

for Table 4.9

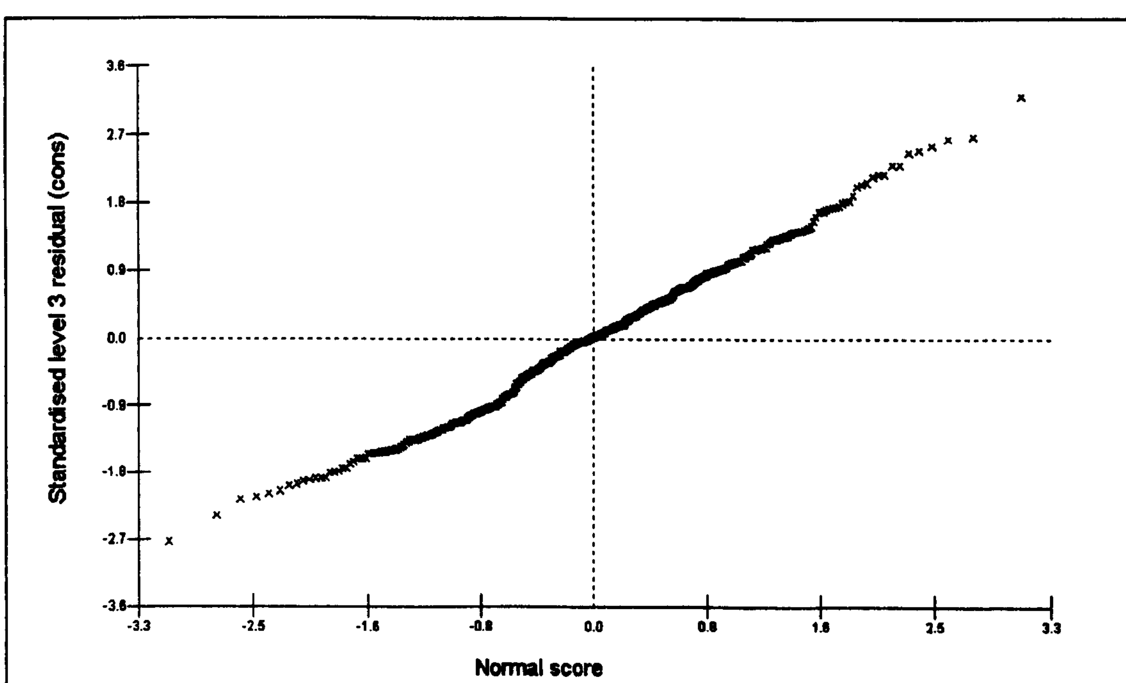


Table 4.9 Parameter estimates after identifying 'abandoned' module entries

Parameter	Estimate	Standard error
Fixed:		
Constant	55.38	0.6439
Project	-2.662	0.275
Male	-0.9945	0.5069
International	-2.016	0.6531
Mature	3.437	0.6413
Coffer	-0.123	0.05555
Professional/managerial	0.1547	0.4816
Workload: 5+	-0.405	0.3476
Year1	-0.2503	0.3238
Term1	0.4444	0.282
Term3	-0.7233	0.2725
Year3	0.6118	0.4344
Stage 2 linear	0.7692	0.09684
Assessment method:		
100%csw	2.245	0.4228
Mixed	2.025	0.3767
Subjects:		
vamp	0.5965	0.3981
bms	-2.11	0.4252
business	1.474	0.3741
cms	0.0005572	1.568
con+es	-1.679	0.5878
educ	0.9881	0.5777
engineer	-2.201	1.625
h+rm	0.3409	0.9078
human	-0.815	0.377
lang	-8.452	1.915
planning	1.686	0.4429
law	-1.963	0.828
unattach	-0.1598	0.8941
Double	0.5688	0.1858
Root class size (centred)	-0.218	0.05397
Stage 2 basic	-0.05495	0.5206
Interactions:		
Profman x year3	0.9743	0.3751
Size x year1	0.1149	0.05062
Allcsw x bms	3.243	0.7188
Allcsw x bus	0.4399	0.4455
Allcsw x cms	-2.247	1.639
Allcsw x engineer	8.651	2.806
Allcsw x h+rm	3.033	1.161
Allcsw x human	0.4411	0.4512
Allcsw x lang	12.07	2.4
Allcsw x law	0.7583	0.9239
Mature x somecsw	-0.9888	0.3425
Size x bms	0.2312	0.105
Size x bus	0.317	0.06538
Size x cms	0.1684	0.1247
Size x c+es	-0.4506	0.1949
Size x h+rm	0.7551	0.2252

Size x law	-0.5988	0.1121
male x year1	1.019	0.3735
mature x year3	-0.09782	0.5793
mature x st2linear	-0.3974	0.1594
somecsw x cms	1.605	1.594
some csw x law	-1.499	0.7706
male x somecsw	-0.4201	0.3137
takeout	-34.93	0.7524
Random:		
Student level: σ^2_{vcons}	37.99	4.632
$\sigma_{vcons/allcsw}$	-15.21	3.352
$\sigma^2_{vallcsw}$	13.23	3.04
$\sigma_{vcons/somecsw}$	-8.205	2.855
$\sigma_{vallcsw/somecsw}$	6.757	2.393
$\sigma^2_{vsomecsw}$	4.995	2.245
$\sigma_{vcons/mature}$	6.23	2.395
$\sigma_{vcyear1/allcsw}$	-3.604	0.8474
$\sigma_{vyear1/somecsw}$	-3.103	0.8147
σ^2_{vyear1}	6.512	1.022
σ^2_{vyear3}	7.497	0.9978
Between terms σ^2_u	0	0
Level 1 variance		
σ^2 [illegible]	57.45	3.449
Covariance between level 1 residuals associated with cons and:		
project	-2.741	1.004
100% coursework	-12.31	1.747
mixed assessment	-14.22	1.724
bms	19.69	1.582
bus	10.32	1.081
cms	13.13	2.882
con+es	8.83	1.576
engineer	28.84	8.84
lang	26.2	8.478
law	-0.315	1.611
workload 5+	4.783	1.552
male	4.71	0.6245
mature	-1.473	0.6253
class size	-0.5272	0.155
year1	6.53	0.7929
stage2 basic	11.63	3.125
size xyear1	-0.5612	0.2101
bms x allcsw	-15.16	2.998
bus x allcsw	-8.559	1.653
cms x somecsw	35.09	4.708
law x somecsw	4.147	2.076
-2*log(lh)	99196	

As expected, fitting a separate term identifying 'abandoned' module entries produces changes in the estimates of level 1 variance parameters: likelihood ratio tests showed that several random parameter estimates in Table 4.9 did not contribute significantly to the model. Variables contributing to the model were reviewed, leading to the removal of a number of terms: the parameter estimates for the final model are shown in Table 4.10.

Checking the residuals after streamlining the model showed that both level 1 and level 3 residuals have approximate Normal distribution (see Figures 4.9 and 4.10).

Figure 4.9 Standardised level 1 residuals by equivalent Normal scores for streamlined model

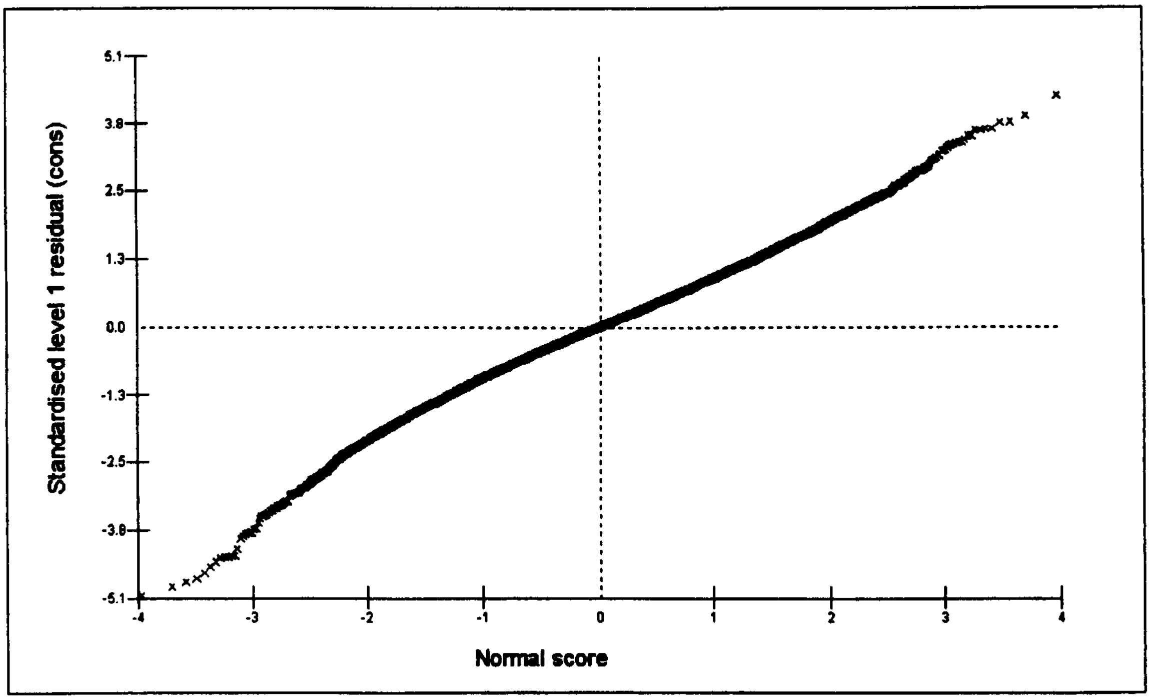
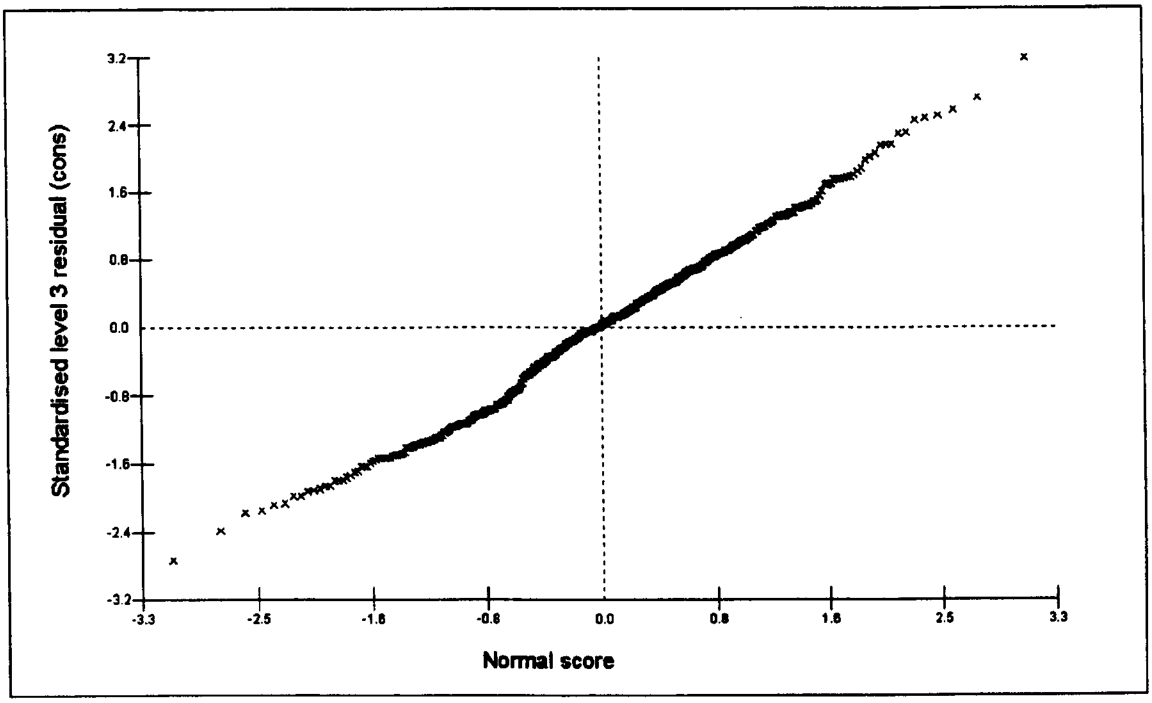


Table 4.10 Parameter estimates after 'streamlining'

Parameter	Estimate	Standard error
Fixed:		
Constant	55.41	0.621
Project	-2.701	0.2695
Male	-1.189	0.4909
International	-2.025	0.6522
Mature	3.467	0.6225
Coffer	-0.1199	0.05512
Professional/managerial	0.1705	0.481
Year1	-0.4079	0.2936
Term3	-0.5321	0.235
Year3	0.543	0.3956
Stage 2 linear	0.7704	0.0917
Assessment method:		
100%csw	2.669	0.352
Mixed	1.977	0.3316
Subjects:		
bms	-2.095	0.4166
business	1.445	0.3228
cms	1.472	0.5302
con+es	-1.911	0.5817
engineer	-2.061	1.62
h+rm	0.278	0.8956
human	-0.9362	0.2794
lang	-8.376	1.904
planning	1.464	0.431
law	-1.776	0.5383
Double	0.6119	0.1817
Root class size (centred)	-0.165	0.04497
Interactions:		
Profman x year3	0.9916	0.3731
Size x year1	0.1198	0.04919
Allcsw x bms	2.251	0.6277
Allcsw x cms	-3.822	0.6945
Allcsw x engineering	8.433	2.803
Allcsw x h+rm	2.585	1.147
Allcsw x lang	11.53	2.399
Mature x somecsw	-1.037	0.3388
Size x bus	0.254	0.05501
Size x c+es	-0.5229	0.1923
Size x h+rm	0.6977	0.2233
Size x law	-0.6805	0.1061
male x year1	0.9766	0.3734
mature x st2linear	-0.4174	0.1148
somecsw x law	-1.7535	0.5567
takeout	-34.94	0.7527
Random:		
Student level: σ^2_{vcons}	35.57	4.612
$\sigma_{vcons/allcsw}$	-15.19	3.358
$\sigma^2_{vallcsw}$	13.54	3.069
$\sigma_{vcons/somecsw}$	-7.953	2.842
$\sigma_{vallcsw/somecsw}$		

$\sigma^2_{vsomecsw}$	6.722	2.4
$\sigma_{vcons/mature}$	4.761	2.235
$\sigma_{vcyear1/allcsw}$	6.329	2.401
$\sigma_{vyear1/somecsw}$	-3.593	0.8524
σ^2_{vyear1}	-3.006	0.8138
σ^2_{vyear3}	6.584	1.026
	7.495	0.998
Between terms σ^2_u	0	0
Level 1 variance		
σ^2_{econs}	57.66	3.398
Covariance between level 1 residuals associated with cons and:		
project	-2.849	0.9958
100% coursework	-12.49	1.74
mixed assessment	-14.32	1.697
bms	19.79	1.587
bus	10.33	1.08
cms	12.83	2.853
con+es	8.772	1.574
engineer	28.96	8.869
lang	26.5	8.538
workload 5+	4.662	1.545
male	4.725	0.6239
mature	-1.441	0.6239
root class size	-0.5461	0.1543
year1	6.571	0.7934
stage2 basic	11.88	3.143
size xyear1	-0.5536	0.2096
bms x allcsw	-15.16	3.0
bus x allcsw	-8.47	1.648
cms x somecsw	35.81	4.711
law x somecsw	3.873	1.339
-2*log(lh)	99213.7	

Figure 4.10 Standardised level 3 residuals by equivalent Normal scores for streamlined model



The estimated residuals were plotted against the predicted responses to check the constant variance assumption. Figures 4.11 shows the estimated level 1 residuals. 'Abandoned' and 'completed' module entries appear as two distinct clusters in Figure 4.11; Figure 4.11a provides a closer look at the estimated residuals for the 'completed' module entries only. Figure 4.12 shows the estimated student-level residuals plotted against predicted student means. These figures support the assumptions of constant variances.

Figure 4.11 Standardised level 1 residuals vs predicted values

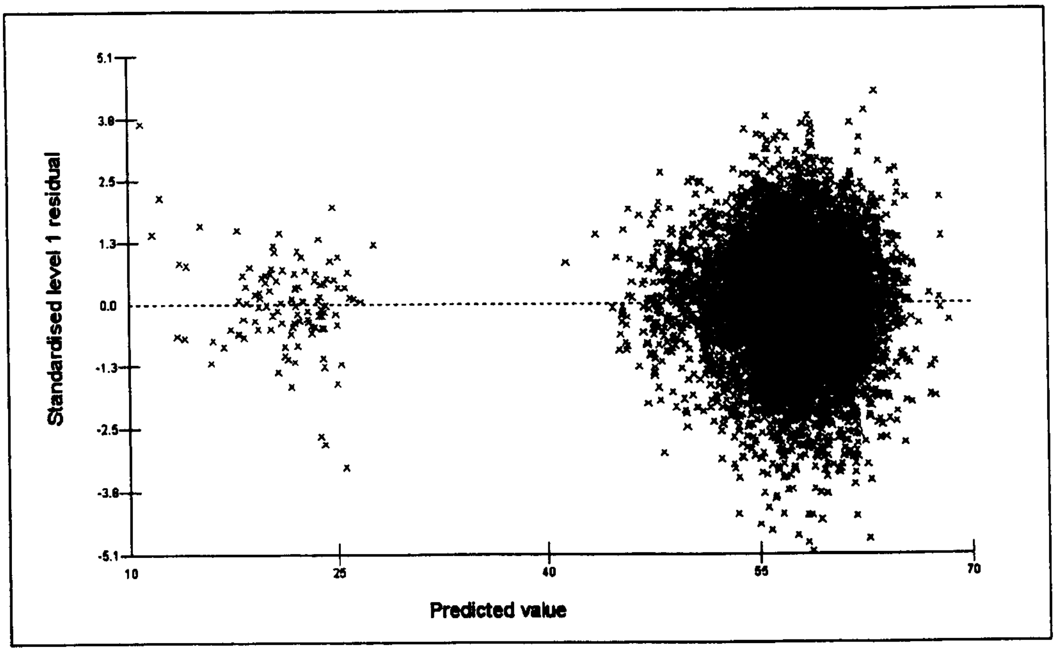


Figure 4.11a Standardised level 1 residuals vs predicted values: completed module entries only

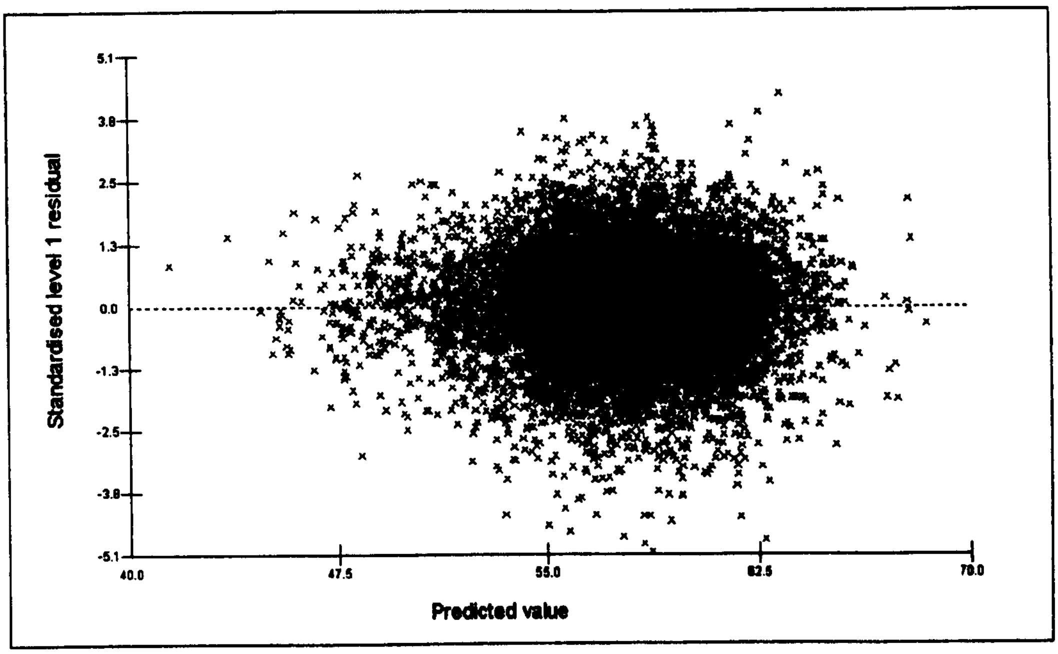
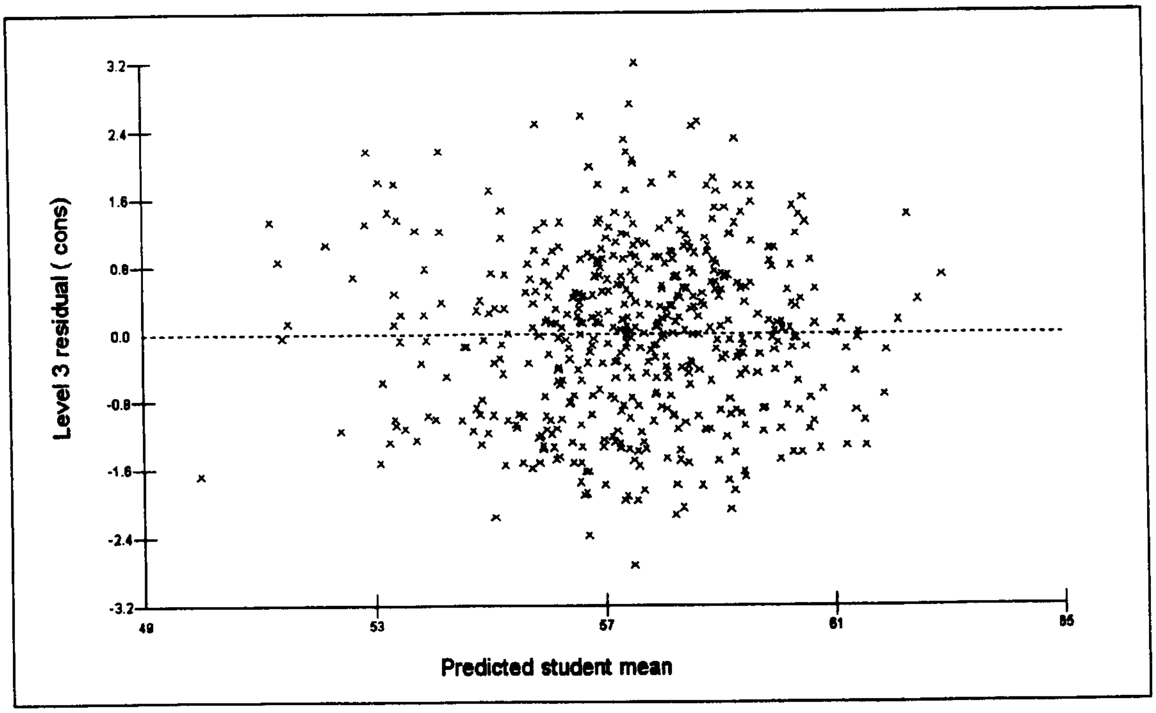


Figure 4.12 Standardised level 3 residuals vs predicted values



The model estimated in Table 4.10, based on equation (4.4), includes all the effects identified as contributing to the model, after streamlining to remove non-significant effects that are not required for the interaction terms. Residual checks support the assumptions of Normally distributed residuals with constant variance.

4.6 Summary

This chapter has explained the decision to use raw marks to measure student achievement and described a model for analysing the student record data based on a 3-level hierarchical structure. Starting with a simple variance components model, the model was developed by adding fixed effects, a complex variance structure at level 1, individual progress charts, random effects and a complex variance term at level 3. A main effects model was used to explore the most appropriate form for representing students' progress: the chosen form described 'progress' as random in the first year and linear afterwards with a 'step' between the second and third years. When the parameters describing progress were allowed to vary between students, the variation between terms was completely explained. The models fitted in this chapter identify factors affecting both the mean levels of attainment and variation in marks, within and between student programmes; variations between students in their patterns of progress and their responses to different forms of assessment have also been identified. These findings will be discussed in detail in chapter 6.

The model was revised after an examination of estimated residuals showed that level 1 residuals were not Normally distributed. Variables were added to the complex variance structure at level 1 and a number of module entries (level 1 units) which may not have been completed in the normal way were identified on the basis of exceptionally low marks. These 'abandoned' entries remain in the analyses but a term was added to the model to distinguish them from other module entries. After these amendments, both level 1 and level 3 estimated residuals appeared to conform to the assumptions underlying the model. At this point a number of non-significant effects were removed from the model, with no adverse effect on the residuals.

All the analyses reported in this chapter are based on a hierarchical model for student record data: the next chapter reviews the appropriateness of this structure.

Chapter 5

Models Based on Alternative Structures

5.1 Introduction

The last chapter introduced a series of models based on a three level hierarchical structure. These models recognised the clustering of individual module entries within the terms of students' degree programmes, but ignored the clustering of module entries made by different students within the same modules. In this chapter, the assumption of a hierarchical structure is reviewed and models based on alternative structures, involving hierarchical and cross-classified elements, are proposed in the next section. Subsets of the first year data are used to fit and compare models based on alternative structures. Data from the first year is used to make these comparisons because the clustering of students' module entries is greatest within the first year, when module enrolments are largest.

The introduction of cross-classification produces changes in the structure of the variance matrix of the responses: with a hierarchical model, this variance matrix has block diagonal structure but, in section 5.3, an example shows that this block diagonal structure is lost when cross-classification is introduced. When the number of cross-classified units is large, using IGLS to obtain parameter estimates becomes difficult. Section 5.4 explores the extent to which the computational demands of such analyses can be reduced by careful specification of the model.

Section 5.5 describes two alternative approaches to fitting multilevel, cross-classified models, both using MCMC estimation. Section 5.6 compares the results achieved and the practical difficulties encountered in using each method of estimation to fit multilevel, cross-classified models to data from the first year.

5.2 Accuracy of Three Level Hierarchical Structure

In the models presented in chapter 4, the assessment data recorded in students' programmes are represented as arising from a simple three level hierarchical structure in which each mark records a students' achievement in one module. To discuss the structure of the model, we will return to the simplest variance components model represented by equation (5.1).

(5.1) $y_{ijk} = \beta_0 + v_k + u_{jk} + e_{ijk}$

where y_{ijk} = mark achieved in the i^{th} module taken by student k in term j

β_0 is a fixed parameter and v_k, u_{jk}, e_{ijk} are the student, term and module level residuals. These are independently distributed, with $v_k \sim N(0,\sigma_v^2)$, $u_{jk} \sim N(0,\sigma_u^2)$ and $e_{ijk} \sim N(0,\sigma_e^2)$.

In fact, the structure of the data is more complex. As all the students' programmes consist of modules chosen from those available within the Modular Degree Programme, many modules were taken by more than one student in the sample. This means that students are crossed with modules, introducing a fourth set of units (modules) and creating a new structure with both hierarchical and cross-classified elements. Whereas in model (5.1), the marks achieved by different students are independent, in reality there will be correlations between the marks achieved by different students in the same module.

The precise details of the new structure depend on how 'modules' are defined: the majority of modules included in students' programmes will have been taught on more than one occasion during the three year period covered by the data. Some students with credits for the same module may have taken it on different occasions. Different 'runs' of a module may be regarded either as different units or as repetitions of the same unit or module. In this section, models based on each of these definitions will be considered.

There is some justification for regarding successive runs of a module as different units, as students taking a module on different occasions have attended different lectures and classes in the company of different peer groups. There may also be changes of teaching staff, examiners and other resources from one run of a module to another. New examinations are set on each occasion and successive runs of a module are considered by the relevant examination committees on separate occasions. If module repetitions are regarded as different units, then students, the level 3 units, are crossed with modules, and the structure of the data is as shown in Figure 5.1. This diagram follows Browne et al. (2001), using arrows to indicate membership of higher level units.

The cross-classification of students by module repetitions means that level 1 units are clustered within two sets of units: one set of units (module repetitions) defines a single level while the other, consisting of terms nested within students, has a hierarchical structure. Figure 5.2 shows how responses might be generated within this structure. For the student record data analysed here, the majority of cells in the cross-classification table are empty : a student might include 30 modules in their programme but the student records for the whole sample cover 1544 module repetitions. Also, as module repetitions are unrelated, each cell contains at most one response, since a student cannot take a module more than once on the same occasion.

Figure 5.1 Cross-classified model: assuming module repetitions are independent units

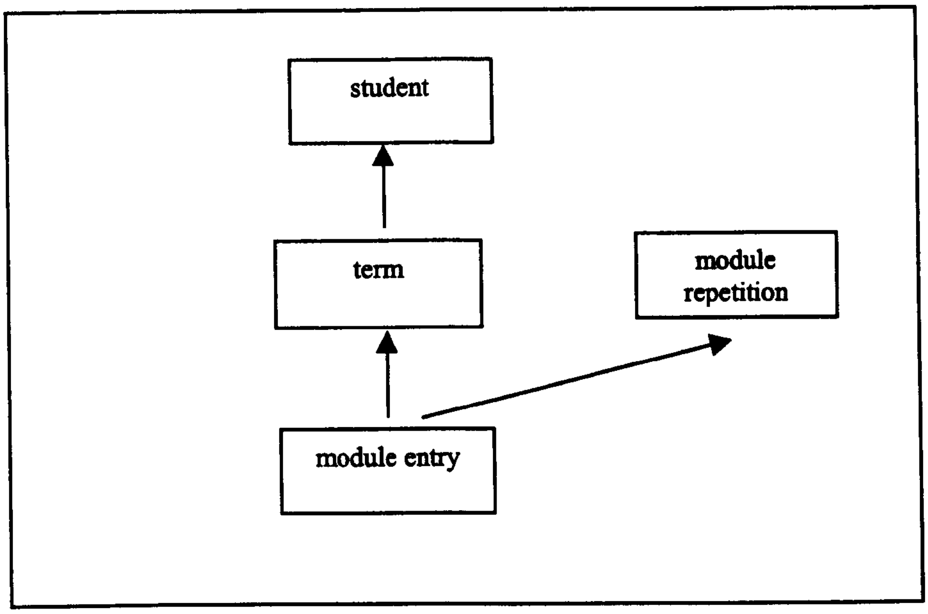


Figure 5.2 Responses generated by structure shown in Figure 5.1

			students			
		1	2	3	.	N_S
		123456789	123456789	123456789	.	123456789
	1	X	X			
Module	2	X				
repetitions	3	X				
	.		X			
	.			X		X
	N_M	X				

The corresponding variance components model is:

$$(5.2) \quad y_{ij(k_1k_2)} = \beta_0 + v_{k_1}^{(1)} + v_{k_2}^{(2)} + u_{jk_1} + e_{ij(k_1k_2)}$$

Individual module entries are indexed by the subscript i, terms are indexed by the subscript j and students and modules are indexed by the subscripts k_1 and k_2

respectively. Subscripts relating to sets of units which are crossed are grouped within brackets.

u_{jk_1} and $e_{ij(k_1k_2)}$ are the level 1 and level 2 residuals, independently distributed with $u_{jk_1} \sim N(0,\sigma_u^2)$ and $e_{ij(k_1k_2)} \sim N(0,\sigma_e^2)$.

$v_{k_1}^{(1)}$ and $v_{k_2}^{(2)}$ are the residuals associated with students and modules respectively, independently distributed with $v_{k_1}^{(1)} \sim N(0,\sigma_{v1}^2)$ and $v_{k_2}^{(2)} \sim N(0,\sigma_{v2}^2)$

For this model, the variation in marks includes four terms, representing variation at level 1, between students, between modules and between terms within student programmes, so that

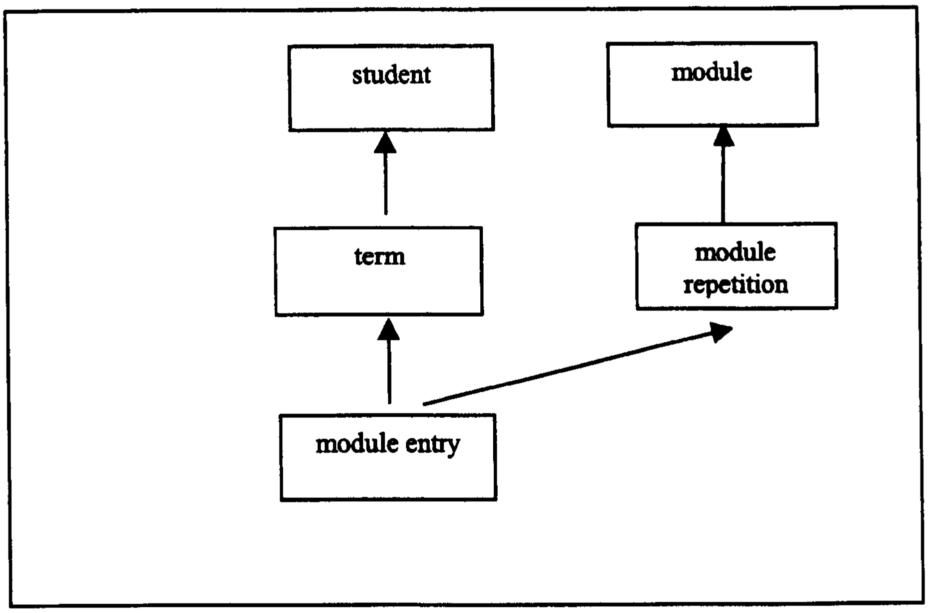
$$\mathrm{var}(y_{ij(k_1k_2)}) = \sigma_{v1}^2 + \sigma_{v2}^2 + \sigma_u^2 + \sigma_e^2$$

As for model (5.1), there are relationships between marks achieved by the same student and, in addition, between marks achieved in the same module:

$$\mathrm{cov}(y_{ij(k_1k_2)}, y_{i'j(k_1k_2')}) = \sigma_{v1}^2 + \sigma_u^2 \text{ (same student and term, different module entries)}$$
$$\mathrm{cov}(y_{ij(k_1k_2)}, y_{i'j'(k_1k_2')}) = \sigma_{v1}^2 \text{ (same student, different terms and modules)}$$
$$\mathrm{cov}(y_{ij(k_1k_2)}, y_{i'j(k_1'k_2)}) = \sigma_{v2}^2 \text{ (same module and term, different students)}$$
$$\mathrm{cov}(y_{ij(k_1k_2)}, y_{i'j'(k_1'k_2')}) = 0 \text{ (any term, different students and different modules)}$$

Although there are reasons for regarding successive runs of a module as unrelated, there are also reasons for adopting the opposite view: a common curriculum, similar assessments and, in many cases, continuity of teaching staff, examiners, teaching materials and other resources. If this view is adopted then repetitions of a module are clustered within modules and the structure of the data is more complicated. This structure, involving 5 sets of units, is shown in Figure 5.3.

Figure 5.3 Cross-classified model with repetitions nested within modules



As in Figure 5.1, the structure has both hierarchical and cross-classified elements: module entries, at level 1, belong to two sets of higher level units, each with their own two level hierarchical structure: terms within students and repetitions within modules. The structure in Figure 5.3 is a combination of two hierarchies: one depicting the assessment system from the students' perspective, with assessments taking place within modules clustered within terms and the other depicting assessment from an examiner's perspective, with students' entries being assessed within each repetition of each module. Each of these hierarchies represents only part of the structure, since the level 1 units, individual module entries, are the same in both cases.

Figure 5.4 shows the responses that might be generated within this structure: as a typical student might take 30 modules, and the programmes for the whole cohort featured 828 different modules, the majority of cells in the table are empty. Some modules feature in the programmes of the sample studied on only occasion, while others appear on more than one occasion. Although most non-empty cells will contain only one response, there will be some cells containing two or more responses, when a student has taken the same module on more than one occasion.

Figure 5.4 Responses generated within structure shown in Figure 5.3

		Student 1	Student 2	Student 3	.	Student N
		Term 123456789	123456789			123456789
Module1	Run1 Run2 Run3	X 	 X 			X
Module2	Run1 Run2	 X		X 		 X
Module3	Run1					

Indexing level 1 units by the subscript i; students by k_1 and terms within student programmes by j_1 and indexing modules by k_2 and repetitions within modules by j_2, then $v_{k_1}^{(1)}, u_{j_1k_1}^{(1)}, v_{k_2}^{(2)}, u_{j_2k_2}^{(2)}, e_{i(j_1k_1,j_2k_2)}$ are random variables introducing variation between students and between terms within student programmes, between modules and between module repetitions within modules, and at level 1 respectively. This can be expressed by the model shown in equation (5.3).

$$(5.3) \quad y_{i(j_1k_1,j_2k_2)} = \beta_0 + v_{k_1}^{(1)} + v_{k_2}^{(2)} + u_{j_1k_1}^{(1)} + u_{j_2k_2}^{(2)} + e_{i(j_1k_1,j_2k_2)}$$

The terms in equation (5.3) are defined as follows:

$e_{i(j_1k_1,j_2k_2)}$ is the level 1 (module entry) residual, independently distributed with $e_{i(j_1k_1,j_2k_2)} \sim N(0,\sigma_e^2)$

Other residuals, also independently distributed, are $v_{k_1}^{(1)} \sim N(0,\sigma_{v1}^2)$, associated with students; $v_{k_2}^{(2)} \sim N(0,\sigma_{v2}^2)$ associated with modules; $u_{j_1k_1} \sim N(0,\sigma_{u1}^2)$ associated with terms within students' programmes and $u_{j_2k_2} \sim N(0,\sigma_{u2}^2)$ associated with repetitions within modules.

With this model, $\operatorname{var}(y_{i(j_1k_1,j_2k_2)}) = \sigma_{v1}^2 + \sigma_{v2}^2 + \sigma_{u1}^2 + \sigma_{u2}^2 + \sigma_e^2$ and

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1k_1,j_2'k_2')}) = \sigma_{v1}^2 + \sigma_{u1}^2$ (same student and term, different module entries)

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1'k_1,j_2'k_2')}) = \sigma_{v1}^2$ (same student, different terms and modules)

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1'k_1,j_2'k_2)}) = \sigma_{v1}^2 + \sigma_{v2}^2$ (same student, different terms and same module)

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1'k_1',j_2k_2)}) = \sigma_{v2}^2 + \sigma_{u2}^2$ (same module and term, different students)

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1'k_1',j_2'k_2)}) = \sigma_{v2}^2$ (same module, different occasions and students)

$\operatorname{cov}(y_{i(j_1k_1,j_2k_2)}, y_{i'(j_1'k_1',j_2'k_2')}) = 0$ (any term, different students and different modules)

5.3 Structure of covariance matrix: an example

The covariance structures implied by models (5.1), (5.2) and (5.3) are illustrated by the following example. For this example, the data consists of the marks achieved by three students over a period of two terms. The first student took three modules in term one and two modules in term two and the second and third students each took four modules in term one and three modules in term two. The modules, identified by the letters a to k , taken by each student, on each occasion are listed in the table below.

Student	Term	Modules taken
1	1	*a, c, f*
	2	*b, g*
2	1	*h, a, d, e*
	2	*i, e, b*
3	1	*j, c, a*
	2	*k, c, d, b*

All three students took module a in term 1 and module b in term 2.

Student 1 took module c in term 1 while student 3 took module c on two occasions, in term 1 and term 2

Students 2 and 3 took module d in terms 1 and 2 respectively

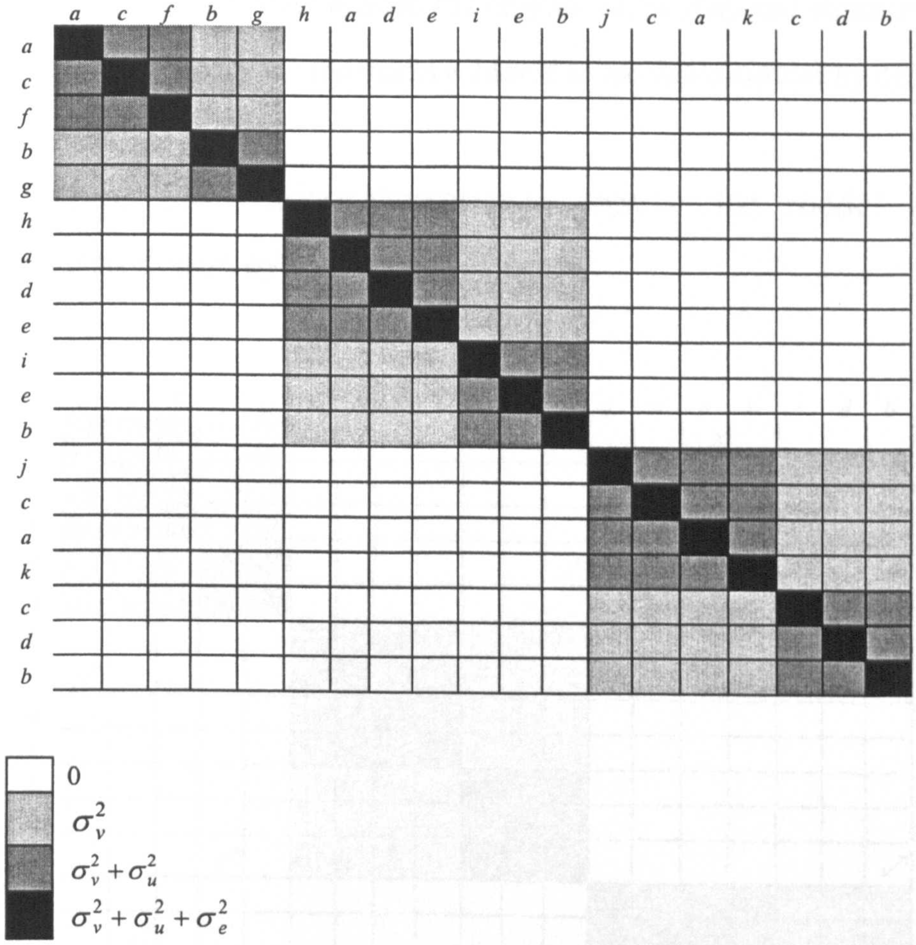
Student 2 takes module *e* in term 1 and repeats it in term 2.

The vector of responses ***y*** is obtained by sorting the responses by student and term.

As the total number of module entries is 19, the variance/covariance matrices for ***y*** implied by each model have dimension 19 x 19. Figures 5.5-5.7 show the structures of the variance/covariance matrices for these models.

With the data sorted by student and term, the variance/covariance matrix associated with model (5.1) has the block diagonal structure shown in Figure 5.5. Using generalised least squares, the estimation requires the inversion of this matrix (Goldstein, 1995, p39). The block diagonal structure means that this inversion will be computationally feasible even if the number of students or modules is large, as was the case in the analyses carried out in chapter 4.

Figure 5.5 Variance/covariance matrix for model (5.1) - 3 level hierarchy



The next diagram shows how the covariance matrix changes when model (5.2) is adopted. The introduction of the cross-classification leads to the appearance of non-zero elements outside the original block diagonal structure, as the result of correlations between the marks achieved in the same module by different students.

Figure 5.6 Variance/covariance matrix for model (5.2) - cross-classification by module repetitions

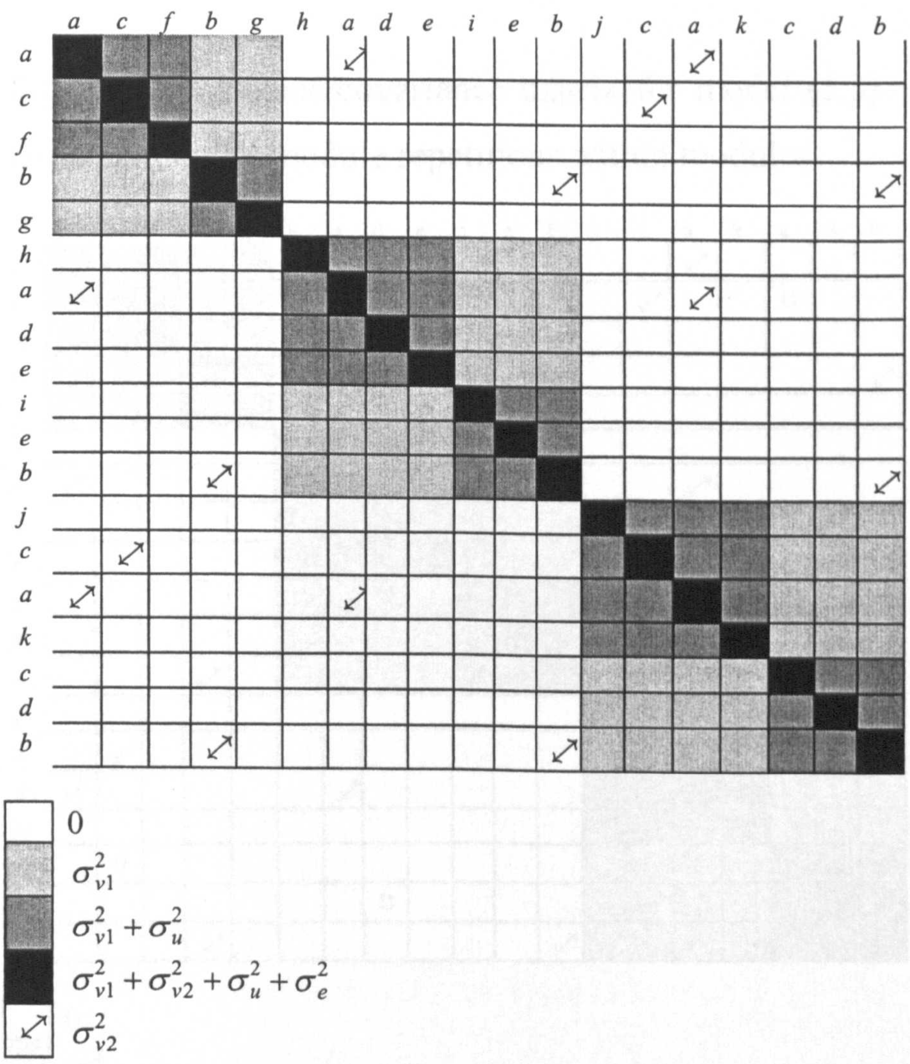
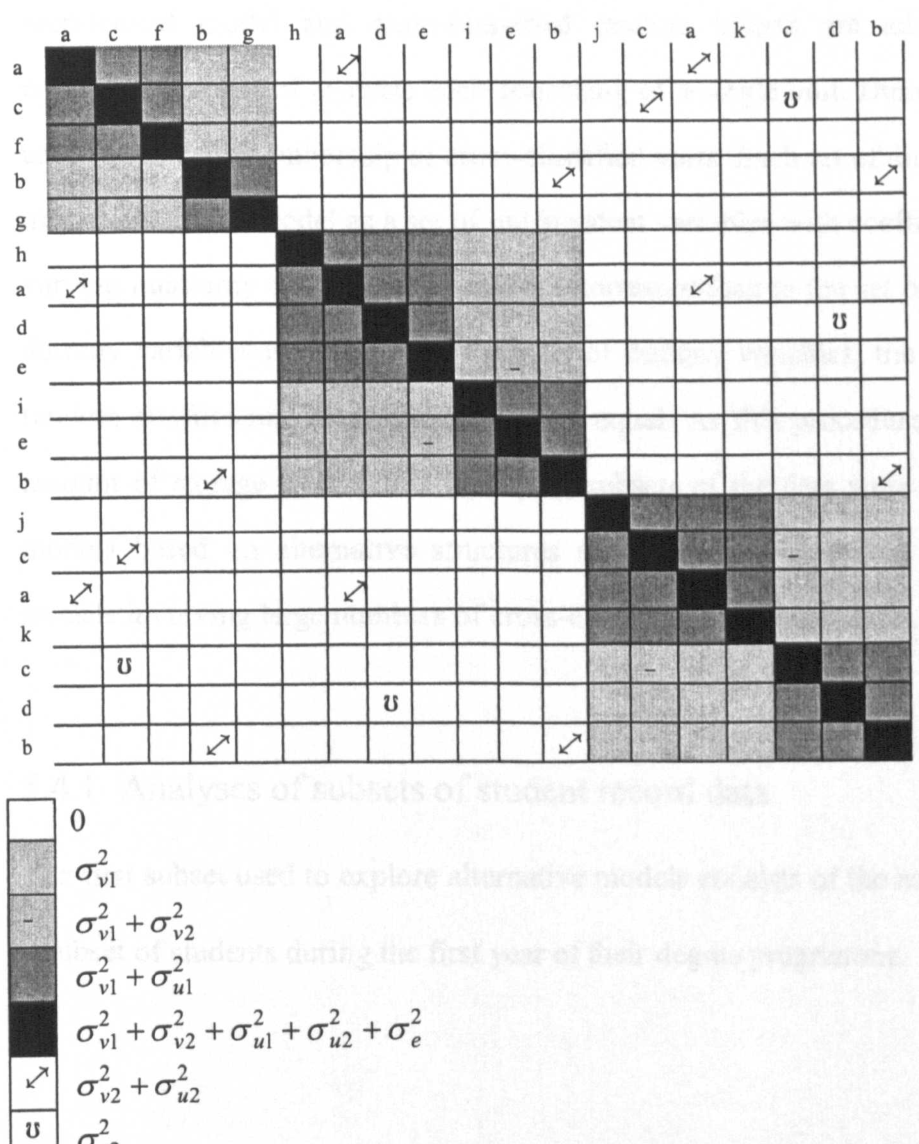


Figure 5.7 shows the variance/covariance matrix associated with model (5.3). As the result of the clustering of repetitions within modules further non-zero elements appear outside the original block diagonal structure. There are also changes to some elements within the original blocks, as the result of correlations between the marks achieved by the same student when taking the same module on different occasions. For both models (5.2) and (5.3), the loss of the block diagonal structure means that inverting the covariance matrix will be computationally demanding for large samples.

Figure 5.7 Variance/covariance matrix for model (5.3) - second cross-classification by module repetitions within modules



The example in this section shows how cross-classification increases the computational demands of fitting a model. The next section uses a subset of the student record data to show how IGLS estimates of the parameters in cross-classified, multilevel models can be obtained using maximum likelihood (Rasbash and Goldstein, 1994) and is followed by a discussion, in section 5.5, of the practical problems associated with fitting these models.

5.4 Fitting cross-classified multilevel models using IGLS

Rasbash and Goldstein (1994) describe a procedure for obtaining maximum likelihood estimates for models involving both hierarchical and crossed random factors. Using this procedure, some random effects are fitted within a simple hierarchical model and cross-classified random effects are added by defining additional level(s) of analysis each consisting of a single unit. Dummy variables are used to indicate membership of cross-classified units. Each set of dummy variables is introduced to the model as a set of independent variables with coefficients defined as varying randomly at the level of analysis corresponding to the set of units which the dummy variables represent. For each set of dummy variables, the variances of the random coefficients are constrained to be equal. As this procedure requires a large amount of storage (Rasbash et al, 2000), subsets of the data were used to compare models based on alternative structures and to explore practical issues in fitting models involving large numbers of cross-classified and nested units.

5.4.1 Analyses of subsets of student record data

The first subset used to explore alternative models consists of the marks achieved by a subset of students during the first year of their degree programme.

This data set includes the following numbers of units:

students	50	modules	123
terms within students	150	repetitions within modules	147
module entries	488		

In this subset relatively few first year modules were taken in different terms by different students. Models (5.1), (5.2) and (5.3) were applied to the marks achieved in the first year by 50 randomly selected students. The parameter estimates for these analyses are shown in Table 5.1. These results show that the models based on more complex, cross-classified structures achieved significant reductions in total deviance when compared to the results based on a simple three-level hierarchical model.

Table 5.1 Results of fitting models (5.1)-(5.3) to subset of first year data

Parameter		Model (5.1) Estimate (se)	Model (5.2) Estimate (se)	Model (5.3) Estimate (se)
Fixed:				
Constant		58.475 (1.002)	58.34 (1.07)	58.45 (1.09)
Random :				
Student	$\sigma^2_{v_1}$	39.274 (10.154)	41.02 (10.27)	42.15 (10.47)
Term	σ_{u_1}	6.103 (5.026)	4.537 (4.521)	3.87 (4.396)
Module	$\sigma^2_{v_2}$		17.01 (5.225)	12.64 (9.008)
Repetition	$\sigma^2_{u_2}$			4.851 (8.884)
Module entry (level 1) σ^2_e		85.369 (6.546)	69.82 (6.279)	69.74 (6.257)
-2logL		3662.743	3645.28	3642.76

Model (5.2) vs Model (5.1) reduction in total deviance = 17.46, df =1
Model (5.3) vs model (5.2) reduction in total deviance = 2.52, df = 1

The second subset of the student record data to be used to compare models based on different structures consists of the marks awarded in term 2 of the first year. A term in the first year provides a good opportunity to study the effects of the clustering of level 1 units within modules as in the first year, basic modules can provide a common foundation for wide variety of programmes. For this term, marks are available for 496 students, in a total of 1,644 module entries, distributed between 80 modules. Only 11 modules were taken by just one member of the sample, confirming the degree of clustering within modules.

For a single occasion, the original three level hierarchy is reduced to a two level model in which module entries are nested within students, so that the variance components model represented by equation (5.1) becomes:

(5.4) $y_{ij} = \beta_0 + u_j + e_{ij}$

where y_{ij} is the mark achieved by student j in the i^{th} module taken on that occasion and u_j and e_{ij} are the student and module entry level residuals, independently distributed with $u_j \sim N(0,\sigma_u^2)$ and $e_{ij} \sim N(0,\sigma_e^2)$.

Since only one repetition of a module occurs on a single occasion, models (5.2) and (5.3) are reduced to the same two level model, with module entries cross-classified, at level 2, by students and modules:

(5.5) $y_{i(j_1j_2)} = \beta_0 + u_{j_1}^{(1)} + u_{j_2}^{(2)} + e_{i(j_1j_2)}$

$e_{i(j_1j_2)}$ is the module entry (level 1) residual and the level 2 residuals, $u_{j_1}^{(1)}$ and $u_{j_2}^{(2)}$ are associated with students and modules respectively. These are independently distributed with $e_{i(j_1j_2)} \sim N(0,\sigma_e^2)$, $u_{j_1} \sim N(0,\sigma_{u1}^2)$ and $u_{j_2} \sim N(0,\sigma_{u2}^2)$

The results of applying models (5.4) and (5.5) to the marks achieved by students in their second term are shown in Table 5.2. Fitting the cross-classification reduces the total deviance by 139.8 with the loss of only 1 degree of freedom, hence the more complex model fits the data significantly better ($P < 0.001$). When variation between modules is not included explicitly in the model, it contributes to the variation at level 1, so that the variability in the performance of students in different modules taken in the same term is overestimated, as the estimate of σ_e^2 rises from 79.85 to 95.53. This was also the case for the analyses shown in Table 5.1.

Table 5.2 illustrates one of the effects of ignoring clustering of responses within modules, that is, the underestimation of the standard errors of parameters. In this example, the estimated standard error of the intercept term is lower when estimates are based on the hierarchical model.

Table 5.2 Results of fitting models (5.4) and (5.5) to data for term 2

	Model (5.4) Estimate (se)	Model (5.5) Estimate (se)
Fixed effects: Constant	57.40 (0.372)	57.68 (0.663)
Random: Level 2 Students ($\sigma_{u_1}^2$) Modules ($\sigma_{u_2}^2$)	39.37 (4.527)	39.1 (4.226) 18.93 (4.313)
Level 1 Module entries (σ_e^2)	95.53 (3.965)	79.85 (3.414)
-2logL	12698.1	12468.3

The results shown in Tables 5.1 and 5.2 confirm that models based on a multilevel, cross-classified structure are more appropriate than a model based on a three level hierarchy. This means that the last model fitted in chapter 4 needs to be updated to include the variation between modules.

Comparing models based on alternative cross-classified structures, the results are less clear cut, and being based on a fraction of cases (488 module entries out of

14,315) and need to be treated with caution. The question of whether different runs of the same module should be treated as separate units will be explored in more detail in the next section. Meanwhile, the next section discusses the practical problems of using IGLS to fit cross-classified multilevel models to all of the student record data.

5.4.2 Practical issues in estimation for multilevel cross-classified models

Finding maximum likelihood estimates of parameters in multilevel, cross-classified models (Rasbash and Goldstein, 1994) can be computationally demanding, requiring a large amount of workspace and taking a long time to converge. Rasbash et al (2000) provide the following formula for calculating the storage required (in worksheet cells) for estimation of parameters in a cross-classified model:

$$3n_e b + n_f b + \sum_{l=1}^{L} 4r_l b + 2br_{\max} \tag{5.6}$$

where n_e = number of explanatory variables

b = number of level 1 units in largest highest level unit

n_f = number of fixed parameters

L = number of levels

r_l = number of variances estimated at level l

$r_{\max}$ = maximum number of variances estimated at a single level

The storage requirements for the analyses reported in Table 5.1 are shown in Table 5.3.

Table 5.3 Worksheet size required to fit models (5.1), (5.2) and (5.3) to subset of first year data

Worksheet cells	Model (5.1)	Model (5.2)	Model (5.3) First specification	Model (5.3) Second specification
First year subset	702	656 x 10^3	1073 x 10^3	837 x 10^3
Running time	3 seconds	23 seconds	-	1110 seconds

Using a Pentium computer with 400MHz and 128MB memory, the 3 level hierarchical variance components model, (5.1), required 702 worksheet cells and convergence was achieved within 3 seconds and 3 iterations. Model (5.2), which incorporates variation between module repetitions, was fitted by defining a fourth level, consisting of a single unit, and creating 147 dummy variables to identify module repetitions. These dummy variables were defined as having coefficients varying randomly at level 4 and constraints used to define their variance as constant. This raised the storage requirements to 656k cells and required a further 3 iterations and a further 20 seconds to converge.

Two attempts were made to fit model (5.3), in which the module repetitions featured in model (5.2) are nested within modules. The first attempt started with model (5.2) specified as above and defined a fifth level, consisting of a single unit and a second set of dummy variables to identify modules. This required 1,073k cells, and increased the number of dummy variables required to fit the cross-classification from 147 to 270. It was not possible to obtain convergence with the model specified in this way. A second attempt to apply model (5.3) was more successful: this time levels 1-3 were used to represent the entries within module repetitions within modules hierarchy and levels 4 and 5 to introduce variation between students and between terms within a student programme. With the model specified in this way, both the number of dummy variables (200), and the storage requirement of 837k cells, are lower than those required for the first specification of model (5.3). Starting with the 'module entries within module repetitions within modules' hierarchy, an attempt to add the variation between student and between terms within programmes simultaneously had not converged after 30 minutes and more than 100 iterations. Building the model in stages was more successful, taking 1.5 minutes and 13 iterations to fit an interim model that incorporated the variation between terms within students' programmes and required 667k cells. The nesting of terms within students was then added by defining a single unit at level 5 and using dummy variables to identify students. This step, taking 17 minutes, was completed within 8 iterations.

The analyses of subsets show that while cross-classified models are more appropriate, the computational demands of fitting these models lead to practical difficulties. The next section is concerned with how to reduce the computational demands of fitting the multilevel, cross-classified models.

5.4.3 Reducing computational demands of multilevel cross-classified models

This section explores three strategies for reducing the computational demands of using IGLS to fit models (5.2) and (5.3), with the objective of being able to apply these models to a larger proportion of the data, and preferably the entire dataset. The strategies are: the choice of which sets of units to fit hierarchically, the identification of separable sub-groups of units within the sample and the use of independent variables to explain variation between one set of units, allowing the removal of one level of variation from the model.

We have already seen, in fitting model (5.3) to the subset of data from the first year, that some reductions in the computational demands can be made by careful specification of the model. Table 5.4 shows the reductions in worksheet requirements for fitting models (5.2) and (5.3) to the whole sample that can be achieved by specifying the model in the most efficient way. In 'specification A' dummy variables are used to represent module repetitions and modules, if necessary, while in 'specification B' dummy variables are used to represent students and terms. 'Specification A' has lower worksheet requirements for both cross-classified models.

Table 5.4 Worksheet requirements for fitting models (5.1), (5.2) and (5.3) to full data

	Model (5.1)	Model (5.2)	Model (5.3)
Variance components:			
Specification A	702	177×10^6	282×10^6
Specification B	702	623×10^6	623×10^6

A second method for reducing worksheet requirements is based on the identification of separable groups of units. When fitting multilevel cross-classified models by IGLS, a major contributor to the number of cells required is b, the size of the largest unit and a common factor of all four terms in (5.6). In the analyses discussed in the previous section, the highest level(s) specified consisted of a single unit and hence b was equal to the total number of level 1 units. The value of b and hence the number of cells required, can be reduced if subgroups of units can be identified such that there are no students or modules shared between sub-groups. If such groups can be found, the highest level can then be defined as consisting of these groups and the value of b is reduced to the number of level 1 units in the largest group (Rasbash et al, 2000). No such groups could be identified in either the subset of first year data or the full dataset. This may be a consequence of the high proportion of students in the sample taking combined honours degrees, the large number of fields which can be combined and the degree of choice allowed within first year programmes.

The dummy variables used to indicate membership of a set of units increase the number of explanatory variables, n_e, the number of variances estimated at levels corresponding to the cross-classified sets of units, r_l , and the maximum number of variances estimated at a single level, r_{max}. All of these increase the number of worksheet cells required, so if one or more sets of units can be removed from the structure on which the model is based, this will lead to reductions in the computational demands of fitting the model. This could be achieved if variation within one set of units can be explained in terms of the independent variables. In the last chapter, variation between terms was explained by 'individualising' students' progress from one year to the next. This allows model (5.3) to be redefined in terms of 4 sets of units rather than 5: module entries (level 1), students and module repetitions nested within modules. If the 'module entries within module repetitions within modules' hierarchy is fitted, then the cross-classified part of the model can be incorporated by defining 496 dummy variables to identify students. This appears to be an improvement, since the original model would require either 4946 dummy variables (using additional levels to incorporate students and terms) or 2672 dummy variables (using additional levels to incorporate modules and repetitions). Unfortunately, as the variation between terms is explained as the result of defining the coefficients of 'year1' and 'year3' as varying randomly at level 3, three parameters are required to describe the variation between students and so the contribution of this part of the model to n_e, r_l and $r_{\max}$ is increased by a factor of 3. The fixed effects used in analysis presented in Table 4.5 increase the size of the worksheet required to fit the hierarchical model to the full sample to 63k cells. Adding the cross-classification by module repetitions and modules further increases the required workspace to 133 x 10^6.

As none of the strategies for reducing workspace requirements enable the cross-classified models to be applied to the full data set using IGLS, alternative methods of estimation based on MCMC and data augmentation were considered. These are described in the next section.

5.5 MCMC Estimation Methods

5.5.1 Introduction to MCMC

This section introduces Markov Chain Monte Carlo (MCMC) methods and briefly outlines how these methods can be applied to estimation of cross-classified, multilevel models. Developments in the use of MCMC methods by Bayesian statisticians are described in detail by Gilks et al (1996), on which this section draws. Stated simply, the Bayesian approach treats unknown model parameters as random variables, using probability distributions to represent knowledge and uncertainty of the parameters' values. A prior distribution, $\pi(\boldsymbol{\theta})$, is used to express existing knowledge of a parameter vector, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots \theta_k)$ and after observing data D, with sampling distribution $p(D|\boldsymbol{\theta})$, Bayes' theorem is used to obtain the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ which encapsulates knowledge of $\boldsymbol{\theta}$ conditional on having observed D. The formula for the posterior distribution is:

$$\pi(\theta|D) = \frac{p(D|\theta)\pi(\theta)}{\int p(D|\theta)\pi(\theta)d\theta} \tag{5.7}$$

Once the posterior distribution has been obtained, it is used to draw conclusions or make decisions based on the expected values of functions, such as $f(\theta)$, chosen as appropriate in the context of the analysis. The expected value of $f(\theta)$, with respect to the posterior distribution, is:

$$E_{\theta|D}[f(\theta)] = \int f(\theta)\pi(\theta|D)d\theta \tag{5.8}$$

In many cases, a suitably chosen prior distribution will lead a Bayesian statistician to obtain the same conclusions as a statistician using frequentist methods, although the language in which these conclusions are expressed and the underlying philosophies will be different. Carlin and Louis (1996) suggest that differences in philosophy can be put aside, especially since in some cases the results produced by Bayesian methods are able to meet the objectives of frequentist statisticians. Similarly, Gilks et al (1996) suggest that as new developments enable Bayesian statisticians to tackle problems which are intractable using conventional frequentist methods, Bayesian methods will become attractive to all applied statisticians. Although the use of a prior distribution is essential, 'uninformative' prior distributions, chosen to express as little prior knowledge of θ as possible, exert minimal influence on the posterior distribution. This can be achieved by choosing values for the parameters of the prior distribution, $\pi(\theta)$, which will make the distribution as flat or 'diffuse' as possible. Taken to extremes, this may lead to the adoption of an 'improper' prior, which does not satisfy the conditions for a valid probability density function. This is acceptable as long as a valid posterior distribution is obtained. The use of uninformative priors overcomes one potential obstacle to acceptance of Bayesian methods by frequentist statisticians.

In practice, integral (5.8) may be difficult to evaluate, particularly for large k, when the integration will be over a large number of parameters. The evaluation of integral (5.8) can be avoided by using Monte Carlo integration in which a sample of n values, $\theta^{(1)},\theta^{(2)},...\theta^{(n)}$, are drawn independently and at random from the posterior distribution $\pi(\theta|D)$ and the average of $f(\theta)$ for these values used to approximate $E_{\theta|D}[f(\theta)]$. In some cases, the posterior distribution (5.7) may itself be difficult to evaluate because of the integral in the denominator, or may be difficult to sample. Using MCMC, these difficulties are avoided by constructing a Markov chain consisting of random draws, $\theta^{(t)}$, from a more tractable distribution which converges, as $t \to \infty$ to the posterior distribution, $\pi(\theta|D)$. A property of Markov chains is that each value depends on the one before, hence the values in the chain are not independent. As each $\theta^{(t)}$ is generated from the previous one in the series, by sampling the distribution $p(\theta^{(t)}|\theta^{(t-1)})$, a starting value, $\theta^{(0)}$, is needed to initiate the chain. For sufficiently large t, $p(\theta^{(t)}|\theta^{(t-1)})$ will approximate the posterior distribution, $\pi(\theta|D)$, and, from this point onward, sampled values approximate correlated random draws from $\pi(\theta|D)$ and can be used to estimate $E_{\theta|D}[f(\theta)]$. Hence if m values are discarded from the start of the chain (usually referred to as the 'burn in') then $E_{\theta|D}[f(\theta)]$ is estimated by the *ergodic average*:

$$\bar{f} = \frac{1}{n-m}\sum_{t=m+1}^{n} f\left(\theta^{(t)}\right) \qquad (5.9)$$

Using equation (5.9), estimates of any function of the model parameters can be obtained from the chain after a suitable burn-in. In particular, sampled values obtained after convergence, are used to estimate the mean and variance of the marginal posterior distribution of each parameter (Gilks et al, 1996, p15):

$$\overline{\theta}_i = \frac{1}{n-m}\sum_{t=m+1}^{n}\theta_i^{(t)} \quad \text{and} \quad s_i^2 = \frac{1}{n-m-1}\sum_{t=m+1}^{n}\left(\theta_i^{(t)} - \overline{\theta}_i\right)^2 \qquad (5.10)$$

To perform MCMC estimation, a number of choices need to be made: the prior distribution for θ, the number of iterations needed overall, n, the length of the 'burn in', m, a starting value, $\theta^{(0)}$ and the selection of an appropriate algorithm for updating the value of θ at each step.

One method for generating a new value for θ from the previous one is the Gibbs sampler (Gilks et al, 1996). This avoids having to sample the joint distribution of a potentially large number of unknown parameters, and hence simplifies the computation involved at each step. When MCMC is implemented using Gibbs sampling, then at each iteration, values are drawn from the full conditional distributions of each element of θ in turn. This means that, at the t^{th} iteration, the value $\theta_i^{(t)}$ $(1 \le i \le k)$ is drawn from the full conditional distribution $\pi\left(\theta_i \middle| D, \theta_1, .. \theta_k \text{ excluding } \theta_i\right)$ with values of $\theta_1, .. \theta_k$ excluding θ_i, supplied by the previous iteration. The next section describes the application of Gibbs sampling to a hierarchical model involving cross-classification.

5.5.2 Application of Gibbs sampling to a multilevel cross-classified model

Equation (5.11) represents a multilevel, cross-classified model of student achievement with module entries (indexed by i) classified by student (j_1) and module (j_2):

(5.11) $y_{i(j_1 j_2)} = \beta X_{i(j_1 j_2)} + u_{j_1}^{(1)} + u_{j_2}^{(2)} + e_{i(j_1 j_2)}$

As the variation between terms is ignored, this model is based on the same structure as the single occasion model in equation (5.5): β is a vector of parameters describing fixed effects, $e_{i(j_1 j_2)}$ is the module entry (level 1) residual and the level 2 residuals, $u_{j_1}^{(1)}$ and $u_{j_2}^{(2)}$ are associated with students and modules respectively. These are independently distributed with $e_{i(j_1 j_2)} \sim N(0, \sigma_e^2)$, $u_{j_1} \sim N(0, \sigma_{u1}^2)$ and $u_{j_2} \sim N(0, \sigma_{u2}^2)$.

Using Gibbs sampling to fit equation (5.11) involves taking random draws from each of the following distributions at each iteration, with the previous iteration supplying values for the parameters treated as known:

$\beta_r \mid Y, X, \beta_1, \ldots \beta_p$ (not including β_r), $u_{j_1}^{(1)}, u_{j_2}^{(2)}, \sigma_{u1}^2, \sigma_{u2}^2, \sigma_e^2$ for $(1 \leq r \leq p)$

$\sigma_{u1}^2 \big| Y, X, \beta, u_{j_1}^{(1)}, u_{j_2}^{(2)}, \sigma_{u2}^2, \sigma_e^2$

$\sigma_{u2}^2 \big| Y, X, \beta, u_{j_1}^{(1)}, u_{j_2}^{(2)}, \sigma_{u1}^2, \sigma_e^2$

$\sigma_e^2 \big| Y, X, \beta, u_{j_1}^{(1)}, u_{j_2}^{(2)}, \sigma_{u1}^2, \sigma_{u2}^2$

$u_{j_1}^{(1)} \big| Y, X, \beta, u_{j_2}^{(2)}, \sigma_{u1}^2, \sigma_{u2}^2, \sigma_e^2$

$u_{j_2}^{(2)} \big| Y, X, \beta, u_{j_1}^{(1)}, \sigma_{u1}^2, \sigma_{u2}^2, \sigma_e^2$

For sufficiently large t, the sampled values will approximate random draws from the joint posterior distribution $\pi\left(\beta, \sigma_{u1}^2, \sigma_{u2}^2, \sigma_e^2, u^{(1)}, u^{(2)}\right)$. After a 'burn in' of m iterations, values from the chain are used to estimate the moments of the marginal posterior distributions of each parameter obtained using the formulae in (5.10).

An updated version of MLwiN (Browne, 2000, personal communication) is able to carry out MCMC estimation of multilevel models with cross-classification and has been used here. In cross-classified models such as (5.11), with simple variance components, (improper) uniform prior distributions ($p(\beta_i) \propto 1$) are assumed for the fixed parameters and diffuse inverse Gamma distributions ($p(1/\sigma^2) \sim Gamma(\varepsilon, \varepsilon)$ where ε is very small) for the variances (Rasbash et al, 2000).

For model (5.11), with p independent variables, Gibbs sampling requires random draws from ($n_{students} + n_{modules} + p$) Normal distributions and three inverse Gamma distributions at each of a potentially large number of iterations. Although the number of calculations required is very large, these calculations are less computationally demanding than the inversion of the covariance matrix for this model (which has structure similar to that shown in Figure 5.6). This means that the MCMC approach can provide estimates for cross-classified models when the smaller of $(n_{students}, n_{modules})$ is too large for a solution to be provided using Rasbash and Goldstein's (1994) procedure.

5.5.3 MCMC Analysis Using The Alternating Imputation Posterior Algorithm

Clayton and Rasbash (1999) developed an alternative MCMC approach to the analysis of multilevel, cross-classified models. This method treats the unknown random effects as missing data, using a data augmentation algorithm developed by Tanner and Wong (1987) and discussed by Schafer (1997). 'Data augmentation' involves assuming values for missing data in order to simplify the process of estimation: Clayton and Rasbash (1999) apply this idea to cross-classified random effects, treating random effects as missing data.

At the t^{th} step, values for one set of cross-classified random effects are used to augment the observed data Y. Given values $\left\{\hat{u}_{j_1}^{(1,t-1)}\right\}$ for the random effects associated with students, equation (5.11) reduces to

$$(5.12)\quad y_{i(j_1j_2)} - \hat{u}_{j_1}^{(1,t-1)} = \beta X_{i(j_1j_2)} + u_{j_2}^{(2)} + e_{i(j_1j_2)}$$

On the left hand side of equation (5.12), the values $\{\hat{u}_{j_1}^{(1,t-1)}\}$ act as offsets, while on the right hand side the random structure is hierarchical, so that a conventional and easily applied method of estimation such as IGLS can be used to estimate β, σ_{u1}^2 and σ_e^2 given $\{\hat{u}_{j_1}^{(1,t-1)}\}$. Random draws from the posterior distributions for these parameters produce the sampled values $\beta^{(t,1)}, \sigma_{u1}^{2(t)}$ and $\sigma_e^{2(t,1)}$ and these in turn are used to sample the conditional distributions of the residuals $\{u_{j2}^{(2,t)}\}$. These sampled values, $\{\hat{u}_{j2}^{(2,t)}\}$, are then substituted into (5.11), leading to an alternative simplified model (5.13):

$$(5.13)\quad y_{i(j_1j_2)} - \hat{u}_{j_2}^{(2,t)} = \beta X_{i(j_1j_2)} + u_{j_1}^{(1)} + e_{i(j_1j_2)}$$

which, like (5.12), has a hierarchical structure. As before, parameter estimates are obtained using IGLS and used to generate the sampled parameter values, $\beta^{(t,2)}, \sigma_{u2}^{2(t)}$ and $\sigma_e^{2(t,2)}$. These in turn are used to sample the distributions of the residuals, obtaining sampled values, $\{\hat{u}_{j_2}^{(2,t)}\}$, which provide the offsets for equation (5.12) at the next step.

As the parameters β and σ_e^2 are involved in both sub-models (5.12) and (5.13), two values for each of these parameters are generated at each iteration. Estimates of the posterior moments of the model parameters are found by averaging over a suitable number of iterations, after suitable burn-in.

Clayton and Rasbash (1999) describe this algorithm as a form of Gibbs sampling in which vectors of parameters are visited in turn. An advantage of the data augmentation approach is that it requires fewer iterations than conventional Gibbs sampling (Schafer, 1997). It is suggested (Clayton and Rasbash, 1999) that for large problems, there is no need to sample the distributions of the random parameters, as the differences between sampled values and estimated values will be negligible.

The AIP algorithm was implemented by writing an MlWin macro applying this method to the assessment data for term 2.

5.5.4 Comparisons between estimates generated by IGLS, MCMC and the AIP algorithm

IGLS and MCMC estimation using Gibbs sampling and the AIP algorithm were used to fit model (5.5) to the term 2 data described earlier. MCMC estimation using Gibbs sampling was performed using starting values obtained by ignoring the cross-classification and using IGLS to estimate the model parameters: estimates were obtained following a burn-in of 500 iterations followed by a chain of 5,000 iterations.

The starting values for $\left\{\hat{u}_{j_1}^{(1,t)}\right\}$ used in the AIP algorithm were obtained by using IGLS to fit the second sub model, (5.13), with no offset. The first 500 iterations were discarded and estimates calculated from the next 4,000 iterations. The parameter estimates obtained using each method are shown in Table 5.5. The results achieved using IGLS and MCMC estimation are very similar, but the AIP algorithm has a lower estimate of variation between modules.

Table 5.5 Parameter estimates obtained using IGLS, AIP and MCMC: term 2 data, cross-classification by students and modules, 1644 level 1 units.

	IGLS	AIP	MCMC
Fixed:			
β_0	57.68 (0.662)	57.66 (0.609)	57.69 (0.670)
Random:			
Modules	18.87 (4.17)	13.35 (1.277)	19.67 (4.863)
Students	39.09 (4.221)	38.95 (0.777)	39.248 (4.205)
Level 1	79.87 (3.423)	81.87 (1.765)	80.19 (3.550)

Using the AIP algorithm, the estimated standard errors of the random parameters are smaller than those obtained by other methods. This suggests that the sample size is too small to justify using estimated values of the random parameters in place of values sampled from the conditional distributions. To see whether the complete dataset is large enough to avoid having to sample random parameters, model (5.5) was fitted to the data for all nine terms using MCMC estimation and the AIP algorithm. The starting values, burn in and chain lengths were the same as before. The results of these analyses are shown in Table 5.6. Note that this sample is too large for IGLS estimation of the cross-classified model.

Table 5.6 Parameter estimates obtained using AIP and MCMC: terms 1-9, cross-classification by students and modules,

14,315 level 1 units, 496 students, 1644 module repetitions

	MCMC	AIP
Fixed: β_0	57.804 (0.287)	57.799 (0.294)
Random:		
Modules	17.801 (1.096)	14.511 (0.289)
Students	32.767 (2.207)	33.678 (0.264)
Level 1	59.630 (0.75)	60.530 (1.096)

In Table 5.6, there is greater agreement between the results produced by the two methods than in Table 5.5, but the AIP algorithm produces a smaller estimate of the variation between modules and lower estimated standard errors for the variances, except at level 1. This suggests that although the number of level 1 units is large enough for estimated rather than sampled values of σ_e^2 to be used, the numbers of units available at other levels are too small, so that without sampling the distributions of $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ at each step, the uncertainty associated with these parameters is underestimated. Using the AIP algorithm to fit a cross-classified model to the complete data set would require the sampling of variance parameters. Given these results, cross-classified analyses of the full data set were carried out using the advanced MCMC estimation procedures available in an update to MLWin (Browne, 2000). The results of these analyses are presented in the next section.

5.5.5 MCMC Estimation for cross-classified model with complex variance structure

In the last model fitted in chapter 4, variation between terms, at level 2, was completely explained, so that the model involved variation between only two sets of units. Adding cross-classification by modules to this model leads to equation (5.14):

(5.14) $y_{i(jk)} = X_{i(jk)}\beta + Z_{i(jk)}\mathbf{v}_k + u_j + W_{i(jk)}\mathbf{e}_{i(jk)}$

where module entries are indexed by *i*, modules by *j* and students by *k*. There are two levels: module entries are level 1 units and students and modules are cross-classified at level 2. Although both students and modules are defined at level 2, in implementing the model, students will be treated nominally as level 3 units, and modules as level 2 units. In line with the implementation of the model, students and modules are indexed by *(j, k)* in equation (5.14) rather than (j_1, j_2): an advantage is that the same notation is used for student-level variance parameters in this model as in earlier models.

ß is a vector of parameters describing fixed effects, $\mathbf{e}_{i(jk)}$ is the module entry (level 1) residual and $\mathbf{v}_k$ and u_j are the residuals associated with students and modules respectively. *X, Z* and *W* are design matrices associated with fixed effects and with random effects associated with students and module entries respectively. u_j has Normal distribution with zero mean and variance σ_u^2, $\mathbf{v}_k$ and $\mathbf{e}_{i(jk)}$ are random vectors, whose elements are Normally distributed, with $E[\mathbf{v}_k]=\mathbf{0}$, $E[\mathbf{v}_k\mathbf{v}_k']=\Omega_v$, $E[\mathbf{e}_{i(jk)}]=\mathbf{0}$ and $E[\mathbf{e}_{i(jk)}\mathbf{e}_{i(jk)}']=\Omega_e$. Note that in this context, repetitions of a module are treated as unrelated.

The prior distributions and the method used to generate estimates at each step are shown in Table 5.7. For the fixed parameters, β, and the level 2 variance, σ_u^2, the prior distributions and method for updating values in the chain are the same as in the MCMC analyses reported in section 5.4.

Model (5.14) has a more complex variance structure than the simple cross-classified models fitted earlier and some changes to the estimation method are required to deal with this. The elements of the variance matrix Ω_v are random parameters representing the random effects of explanatory variables included in the design matrix *Z*. As described in chapter 4, Ω_v is a 6 x 6 matrix of variances and covariances. To estimate the student-level variance parameters, the inverse Wishart distribution replaces the inverse Gamma prior assumed earlier, when there was a single variance parameter at student level. The use of an inverse Wishart distribution means that estimates of Ω_v will form a valid (positive definite) variance matrix. For the final model in chapter 4, Ω_v was defined as follows (lower triangular elements only):

$$\Omega_v = \begin{bmatrix} \sigma^2_{vcons} & & & & & \\ \sigma_{vcons/allcsw} & \sigma^2_{vallcsw} & & & & \\ \sigma_{vcons/somecsw} & \sigma_{vsomecsw/allcsw} & \sigma^2_{vsomecsw} & & & \\ 0 & \sigma_{vyear1/allcsw} & \sigma_{vyear1/somecsw} & \sigma^2_{vyear1} & & \\ 0 & 0 & 0 & 0 & \sigma^2_{vyear3} & \\ \sigma_{vcons/mature} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In this form, Ω_v is not positive definite. Minor changes were made allowing $\sigma_{vcons/year1}$ to be non-zero and using a variance parameter, $\sigma^2_{vmature}$, to represent the impact of student's age on between-student variation instead of the covariance $\sigma_{vcons/mature}$ (whose estimate was positive). This leads to:

$$\Omega_v = \begin{bmatrix} \sigma^2_{vcons} & & & & & \\ \sigma_{vcons/allcsw} & \sigma^2_{vallcsw} & & & & \\ \sigma_{vcons/somecsw} & \sigma_{vsomecsw/allcsw} & \sigma^2_{vsomecsw} & & & \\ \sigma_{vcons/year1} & \sigma_{vyear1/allcsw} & \sigma_{vyear1/somecsw} & \sigma^2_{vyear1} & & \\ 0 & 0 & 0 & 0 & \sigma^2_{vyear3} & \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{vmature} \end{bmatrix}$$

which is positive definite.

Table 5.7 Prior Distributions and Methods for updating Markov Chain

Parameters	Prior	Update method
Fixed ß	Uniform	Gibbs sampling
Random:		
Level 3 Ω_v	Inverse Wishart	Gibbs sampling
Level 2 σ^2_u	Inverse Gamma	Gibbs sampling
Level 1 Ω_e	Joint Uniform	Adaptive Metropolis-Hastings

In model (5.14), the complex variance structure at level 1 is represented by a matrix of parameters, Ω_e. This matrix Ω_e consists largely of zeroes, with non-zero elements consisting of a single variance, σ^2_{econs}, and the covariance terms $\{\sigma_{econs/w_i}, i = 1...20\}$ representing, as explained in section 4.4, the effects of explanatory variables in W on the level 1 variation. Ω_e is not a conventional variance matrix and is not positive definite, so using the invert Wishart distribution to generate estimates that will by definition be positive definite will exclude many possible values of Ω_e. An alternative method for estimating complex level 1 variance parameters is described by Browne et al (2000), imposing the looser constraint that all the composite level 1 residuals $e^*_{i(jk)} = W_{i(jk)}e_{i(jk)}$ must have positive variances. Browne et al (2000) use the Metropolis-Hastings algorithm, an alternative to Gibbs sampling, to update the values of Ω_e. A joint uniform prior distribution is assumed for the level 1 random parameters, and is defined so that all values for which the constraint is satisfied have equal probability and only values for which the constraint is met have non-zero probability. Following the Metropolis-Hastings algorithm, values are updated by sampling a proposal distribution for each random parameter in turn, with values being accepted with probability r or rejected in favour of the value from the previous step.

The proposal distributions for random parameters at level 1 are truncated Normal distributions whose truncation points are chosen to ensure that the constraint on the level 1 variance function is met (Browne, 1998). The variances of the proposal distributions are chosen to obtain a suitable acceptance rate, r, ideally close to 50%. A proposal distribution with too large a variance would lead to widely scattered values being proposed and too many would be rejected, causing the chain to 'stick'. Too small a variance would mean that only small steps were taken with many iterations being required for the chain to move from one area of the parameter space to another and samples to be highly correlated. The software used to implement this method (Browne, 2000) uses an adaptive procedure described by Browne and Draper (2000) to set the variances of the proposal distributions. This procedure, which is iterative, is carried out before commencing the burn-in.

5.6 MCMC Estimation of cross-classified model

Model (5.14) was fitted to the full data set using the method described above: starting values for the fixed parameters and for level 1 and student-level variance parameters were obtained from analyses based on the 'module entries within terms within student programmes' hierarchy. The estimated between-modules variance for the cross-classified variance components model (see Table 5.7) was used as the starting value for σ_u^2. Following the adaptive procedure, a burn in of 500 iterations was followed by a chain of 20,000 iterations. Computation was slow due to the complexity of the model: as 1,000 iterations took 350 minutes, the chain of 20,000 was built up incrementally. The final length of the chain was chosen with regard to the diagnostics which are described in the next chapter.

5.7 Summary and conclusion

This chapter has explored alternatives to the hierarchical structure used in chapter 4 to represent the student record data. Tests based on small samples of data from the first year of students' records showed that models based on a more complex, cross-classified multilevel structure provide a more accurate representation of the structure of the data. Chapter 4 presented estimates of the fixed and random effects of explanatory variables based on a simple hierarchical model; this chapter has shown that a cross-classified hierarchical model is more appropriate and therefore the next objective is to update the model from chapter 4 to include the variation between modules. Parameter estimates for cross-classified multilevel models can be obtained using IGLS but the techniques for doing so cannot be used when large numbers of cross-classified units are involved. Even using all the means available to reduce the computational demands, it was not possible to fit even a variance components model based on a cross-classified model to the full dataset using IGLS.

Two alternative methods of estimation were explored, both using MCMC estimation. The final section of this chapter described how the cross-classification of students by modules can be added to the model fitted in chapter 4, using MCMC estimation as implemented by Browne (2000). The next chapter will present the results of this analysis.

Chapter 6

MCMC Analysis Of Cross Classified Model

6.1 Introduction

This chapter will present the results of using MCMC estimation to fit a model to the student record data. The model includes all the fixed and random effects selected in chapter 4 but will also take into account the clustering of students' module entries within modules. In section 4.5.4 it was found that variation between terms could be completely explained by including certain random effects in the model hence the new structure is a cross-classification of module entries (the level 1 units) by students and modules (both at level 2). The earlier discussion of whether different runs of the same module should be treated as one unit or separate units is resumed in section 6.2.

The final model is restated in section 6.3 and was fitted using the MCMC estimation procedures described in the previous chapter. Section 6.4 presents diagnostic output providing information on the convergence of the parameter estimates. The parameter estimates themselves are presented in section 6.5. The model includes a large number of independent variables which influence performance in a variety of ways. These effects have been grouped according to a number of themes, and are discussed in sections 6.6 onwards.

6.2 Choice of units: modules or module repetitions

Section 5.2 considered whether students taking the same module on different occasions should be regarded as having enrolled in the same or different units. The continuity of some aspects of a module: curriculum, intended learning outcomes, and some aspects of teaching and resources supports the definition of module repetitions as comprising a single unit, while the potential for change in aspects such as assessments, peer group, teaching staff, styles of presentation and the emphasis on different parts of the curriculum supports the treatment of repetitions of a module as different units. In the cross-classified analyses presented in Table 5.1, the estimated variation within modules and between repetitions was small relative to its standard error (4.85 , se = 8.88), however these results were based on a small sample of data from the first year and should therefore be treated with caution. A further difficulty is that, for the cohort of students whose records are analysed here, differences between module repetitions are confounded with the time in a students' programmes. Some of the estimated variation between modules and repetitions would have been due to variation between terms within student programmes since this was not included in the models compared in chapter 5.

Using MCMC estimation, the effects of using different definitions for 'modules' can be explored using all 14.315 module entries and variation between terms can be explicitly included in the model. Table 6.1 shows the parameter estimates obtained using MCMC estimation to fit a variance components model (equation (6.1) below) in which cross-classification by modules is added to the three level student/term/module entry hierarchy.

(6.1) $y_{ij(k_1k_2)} = \beta_0 + v_{k_1}^{(1)} + v_{k_2}^{(2)} + u_{j(k_1k_2)} + e_{ij(k_1k_2)}$

In model (6.1), module entries and terms are indexed by i and j and students and modules by k_1 and k_2 respectively. The residuals $v_{k_1}^{(1)}, v_{k_2}^{(2)}, u_{j(k_1k_2)}$ and $e_{ij(k_1k_2)}$ are independently distributed with $v_{k_1}^{(1)} \sim N(0,\sigma_{v1}^2)$, $v_{k_2}^{(2)} \sim N\left(0,\sigma_{v2}^2\right)$, $u_{j(k_1k_2)} \sim N\left(0,\sigma_u^2\right)$ and $e_{ij(k_1k_2)} \sim N\left(0,\sigma_e^2\right)$.

This model was fitted twice, using each definition of "modules" in turn. Table 6.1 shows that very similar variance estimates were produced at all four levels, however modules are defined.

Table 6.1 Parameter estimates for variance components model: three level hierarchy cross-classified by modules and module repetitions

	Cross classification by:	
	Modules	Module repetitions
Parameter	Estimate (se)	Estimate (se)
Fixed: β_0	57.67 (0.522)	57.60 (0.890)
Random:		
Students σ_{v1}^2	32.95 (2.231)	32.91 (2.250)
Modules σ_{v2}^2	14.36 (1.053)	14.97 (0.974)
Terms σ_u^2	3.17 (2.168)	3.88 (2.773)
Module entries σ_e^2	60.57 (0.759)	59.67 (0.748)

Another issue to consider is the treatment of double modules, in which successful students are awarded two module credits, each carrying an identical mark. Given the way that the outcomes of double modules have been recorded, defining different runs of the same module as separate will lead to the credits awarded in double modules being treated as separate, creating artificial distinctions which will tend to lower the estimated variation between module repetitions. This would tend to support the treatment of repetitions of a module as the same unit.

Although some conditions will change between repetitions of the same module, the intention is for there to be continuity between repetitions and in practice there are likely to be more similarities than differences. In addition, the way that 'modules' are defined appears to have little impact on the estimated parameters in the variance components model. Future analyses will therefore define students enrolling for the same module on different occasions as entering the same unit. The choice of the 'modules' approach also leads to more appropriate treatment of the marks arising from double modules.

6.3 The final model

As this section and those that follow will give detailed information about the parameter estimates obtained by fitting model (5.14) to the student record data using MCMC estimation, at this point it is useful to restate the model and identify individual model parameters. In this model, the variation between terms has been explained so that module entries, the level 1 units, are cross-classified, at level 2, by students (nominally represented as level 3) and modules (represented as level 2). Module entries made within the same module on different occasions are treated as belonging to the same level 2 unit. The model corresponds to model (5.14), given in the last chapter, written as follows:

(6.2) $y_{i(jk)} = X_{i(jk)}\beta + Z_{i(jk)}\mathbf{v}_k + u_j + W_{i(jk)}\mathbf{e}_{i(jk)}$

where module entries are indexed by i, modules by j and students by k. As before, ß is a vector of parameters describing fixed effects, and $\mathbf{v}_k$, u_j and $\mathbf{e}_{i(jk)}$ are the residuals associated with students, modules and module entries respectively.

u_j has Normal distribution with zero mean and variance σ_u^2, $\mathbf{v}_k$ and $\mathbf{e}_{i(jk)}$ are random vectors whose elements are Normally distributed with $E[\mathbf{v}_k] = \mathbf{0}$, $E[\mathbf{v}_k \mathbf{v}_k'] = \Omega_v$, $E[\mathbf{e}_{i(jk)}] = \mathbf{0}$ and $E[e_{i(jk)} e_{i(jk)}'] = \Omega_e$.

X, *Z* and *W* are design matrices associated with fixed effects and with random effects associated with students and module entries respectively.

Equation (6.2) can be rewritten in terms of individual parameters and residuals:

$$(6.3) \quad y_{i(jk)} = \sum_{p=0}^{P} \beta_p x_{pi(jk)} + \sum_{h=0}^{H} z_{hi(jk)} v_{hk} + u_j + \sum_{m=0}^{M} w_{mi(jk)} e_{mi(jk)}$$

$x_0 .. x_P$ are the explanatory variables in the design matrix X: $\{\beta_p, p = 0,1,..P\}$ are the fixed effects of the explanatory variables in X, after allowing for other variables.

In discussing estimates of the β's, variable names will be used as subscripts. For example, x_0 is a dummy variable, equal to 1 for all level 1 units and labelled 'CONS': x_0 has associated parameter β_{cons}.

The parameter β_{cons} represents the expected mark achieved in module entries in a reference category, for which all the explanatory variables (apart from CONS) have value zero. This reference category defines a 'standard' student in a 'standard' module. For the purpose of the analysis, 'standard' students are defined as female and aged under 21 on entry to the course; they have UK domicile, graduate in subjects for which the specified entry qualifications correspond to 15.7 A-level points and do not have a parent in an occupation classified as professional or managerial. A 'standard' module is defined as a single, advanced level module running in the hypothetical term 3.5 of the cohort's second year (term 6.5 of their programme), with a class size of 61.4 , 100% examination assessment and taught and assessed within the School of Social Sciences. It is assumed that the module is completed in the normal way (that is, is not abandoned by the student) and is taken during a term when the student's workload consists of four or fewer modules. Note that the 'standard student' and 'standard module' are not typical of the sample but simply belong to those categories that correspond to the reference category defined above and against which the other categories of students and modules are compared. Student related explanatory variables in X are mainly dummy variables, coded 0 or 1, identifying students belonging to specific categories. An example is the dummy variable 'MALE' coded 1 for male students and 0 for female students; the parameter β_{male} measures the mean difference, other things being equal, between the marks awarded to male students and those awarded to a 'standard' (female) student in a module with the same characteristics. Similarly $\beta_{internat}, \beta_{profman}$ and β_{mature} compare (respectively) the mean marks achieved by international students, students with a parent in a professional or managerial occupation and mature students to those achieved by 'standard students' , other things being equal. Further fixed parameters represent the mean difference between the marks awarded in particular types of module and 'standard modules', after controlling for other factors. The types of module considered and their associated parameters are project modules($\beta_{project}$), double modules (β_{double}), modules using 100% coursework assessment (β_{allcsw}), modules using mixed assessment ($\beta_{somecsw}$) and modules in biology and molecular sciences (β_{bms}), computing and mathematical sciences(β_{cms}), construction and earth sciences (β_{con+es}), hotel and restaurant management (β_{h+rm}), humanities (β_{human}), languages (β_{lang}), planning $(\beta_{planning})$ and law (β_{law}). Fixed parameters describing student's progress are $\beta_{year1}, \beta_{year3}, \beta_{term3}$ and β_{st2lin}: these will be discussed in detail in section 6.7.6.

The dummy variable 'TAKEOUT' identifies module entries which according to criteria described in chapter 4, are believed to have been abandoned by students without completing the assessed work. The parameter $\beta_{takeout}$ is used to fit a separate mean for these module entries as described in the section of chapter 4 dealing with residuals.

$\beta_{crtsize}$ and β_{coffer} are the partial regression coefficients of $\sqrt{\text{class size}}$ and minimum A-level points required for the students chosen fields, both measured about their respective means. There are also a number of interaction terms in the model; these are identified by their subscripts in the same way as main effects and also measure mean marks relative to the reference category described earlier.

In the final model, the vector of student level residuals has elements $\{v_{consk}, v_{allcswk}, v_{somecswk}, v_{year1k}, v_{year3k}, v_{maturek}\}$. The mean level of achievement for (standard) student *k*, in standard modules is $\beta_{cons} + v_{consk}$ and the variation between students in these means is variance σ^2_{vcons}. The difference between the mean marks a (standard) student *k*, achieves in modules using 100% coursework assessment compared to their mean in standard modules is $\beta_{allcsw} + v_{allcswk}$ and the between students variance in these differences is $\sigma^2_{vallcsw}$, and so on. The student level residuals are assumed to be Normally distributed with zero means; their variances and covariances form the diagonal and off-diagonal elements of Ω_v, as shown below:

$$\Omega_v = \begin{bmatrix} \sigma^2_{vcons} & & & & & \\ \sigma_{vcons/allcsw} & \sigma^2_{vallcsw} & & & & \\ \sigma_{vcons/somecsw} & \sigma_{vsomecsw/allcsw} & \sigma^2_{vsomecsw} & & & \\ \sigma_{vcons/year1} & \sigma_{vyear1/allcsw} & \sigma_{vyear1/somecsw} & \sigma^2_{vyear1} & & \\ 0 & 0 & 0 & 0 & \sigma^2_{vyear3} & \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{vmature} \end{bmatrix}$$

The level 1 variance matrix, Ω_e, has 21 non-zero elements: the first, σ^2_{econs}, associated with the dummy variable CONS, is the level 1 variation in the marks achieved by a standard student in standard modules taken in the same term. Other non zero elements of the matrix Ω_e are the covariances between level 1 residuals $\{\sigma_{econs/project}, \sigma_{econs/allcsw},\ldots\sigma_{econs/st2basic}\}$. These represent the effects of the explanatory variables on the level 1 variation, with, for the example, the marks awarded to standard students in a module using 100% coursework assessment having level 1 variation of $\sigma^2_{econs} + 2\sigma_{econs/allcsw}$.

6.4 MCMC Diagnostics

Using updated MLwiN software (Browne, 2000), model (6.2) was fitted to the student record data. This model represents module entries as cross-classified by modules (nominally level 2) and students (nominally level 3) and includes the fixed and random effects and complex variance terms selected in chapter 4, with the addition of variation between modules (σ^2_u) and minor rearrangement of terms in the level 3 variance matrix required for the software to run successfully (Browne et al, 2000). After the adaptive procedure to set the acceptance rate for the Metropolis-

Hastings updating of level 1 variance parameters (Browne and Draper, 2000), and a burn-in of 500 iterations, a chain of 20,000 iterations was generated, building up the length of the chain in increments of 500-2,000 iterations at a time. At a speed of approximately 170 iterations per hour, the computations took 6 days to complete. MCMC diagnostics provided within the MLwiN programme (Rasbash et al, 2000) were used to monitor convergence and determine the final length of the chain.

The diagnostics showed that the estimation of some parameters required more iterations than others. Figures 6.1 and 6.2 show the diagnostic output for two parameters: the fixed parameter β_{cms}, the coefficient for the dummy variable identifying module entries in computing and mathematical sciences (CMS) and the random parameter $\sigma_{vallcsw/somecsw}$, the covariance between the student level residuals associated with the dummy variables identifying modules using mixed assessment methods and modules using 100% coursework assessment.

In Figure 6.1, the graph at the top left plots the entire chain of values for β_{cms}. If the chain has converged, this graph should resemble pure random variation, as is the case here. At the top right of Figure 6.1 is a graph showing the kernel density estimate of the posterior distribution which appears to be approximately Normal.

The second row of output in Figure 6.1 shows the auto-correlation (ACF) and partial auto-correlation (PACF) functions of the chain of values for β_{cms}. The chain appears to behave as a first order auto regressive time series. The PACF has a strong spike at lag 1, with a correlation of approximately 0.75 between adjacent values, while the partial auto-correlations for other time lags are close to zero. The ACF also shows that correlations between values in the chain only become small when values are 10 or more steps apart. In general, the greater the correlation between values in the chain, the longer the chain length required to produce a precise estimate of the parameter.

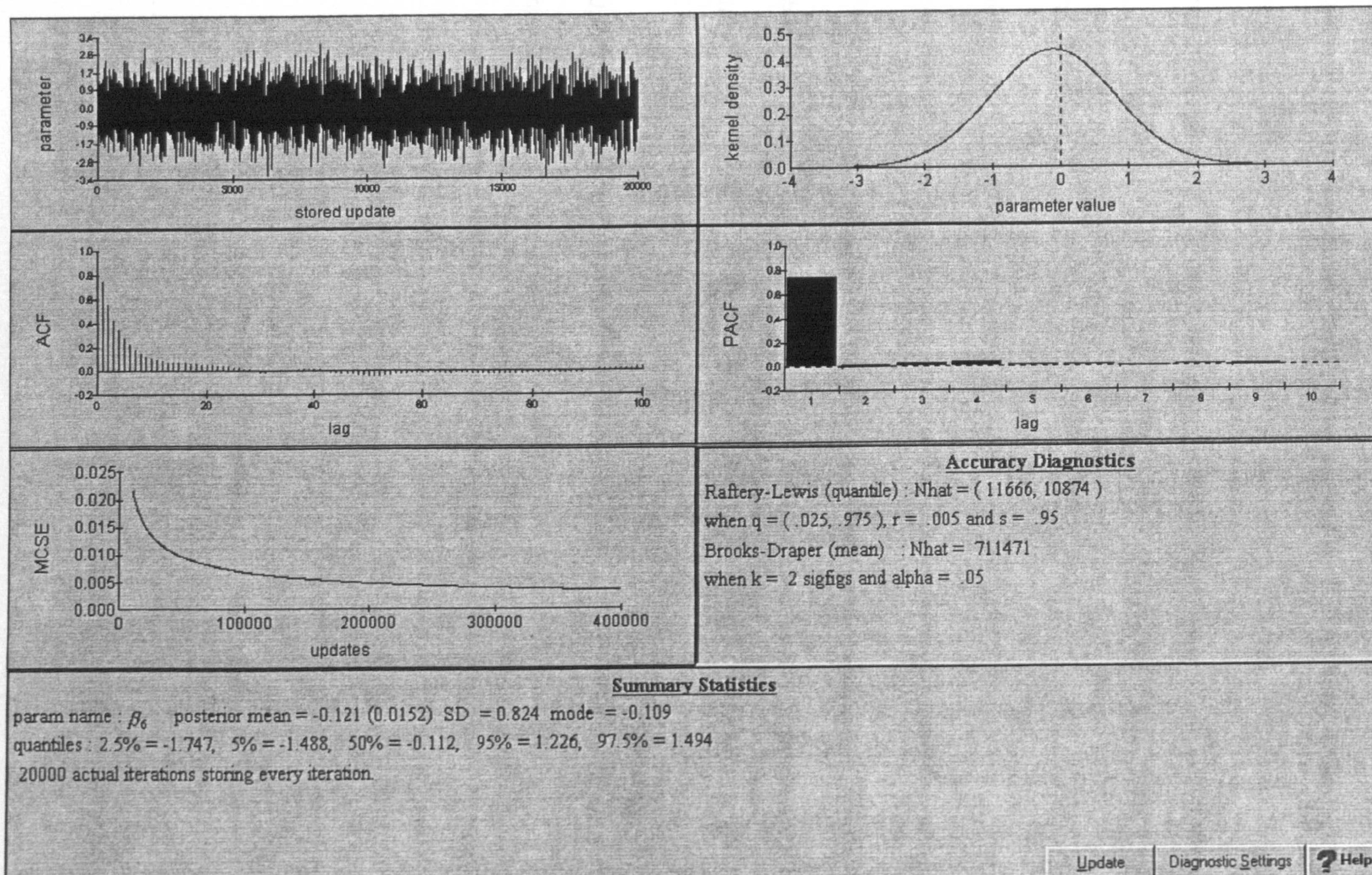
In the third row of output in Figure 6.1, the graph on the left shows the estimated Monte Carlo standard error plotted against the length of the chain. This shows the chain lengths required for the mean of the posterior distribution to be estimated with given Monte Carlo standard error (MCSE). For 20,000 iterations, the MCSE for the mean of β_{cms} is 0.015.

Accuracy diagnostics given on the right hand side of Figure 6.1 provide values of $\hat{N}$ estimating the chain lengths required in order to estimate certain characteristics of the posterior distribution with specified levels of accuracy. Browne (1998) reports that "*in terms of coverage properties, the MCMC methods do better than maximum likelihood in many situations, though not always*". The Raftery-Lewis values of $\hat{N}$ (Raftery and Lewis, 1992) are the estimated chain lengths required to estimate selected percentiles of the posterior distribution, with specified accuracy. In Figure 6.1 the values of $\hat{N}$ are calculated so as to estimate the 2.5% and 97.5% percentiles for β_{cms} to within $\pm 100r$, where $r = 0.005$, with probability $s = 0.95$. These estimated percentiles may then be used to provide confidence intervals with specified coverage of the posterior distribution. For β_6, the Raftery-Lewis calculations show that a chain length of 11,666 will ensure that the 95% confidence interval for β_{cms} will actually cover between 94% and 96% of the distribution with probability 0.95.

The Brooks-Draper value of $\hat{N}$ estimates the chain length required to estimate the mean of the posterior distribution with specified accuracy. To estimate the mean of the posterior distribution of β_{cms} to 2 significant figures would require

711,471 iterations, however, given the order of magnitude of β_{cms}, this is more accurate than necessary. An estimate of the mean of β_{cms} which is accurate to one significant figure is sufficient, and requires a chain length of approximately 7,115.

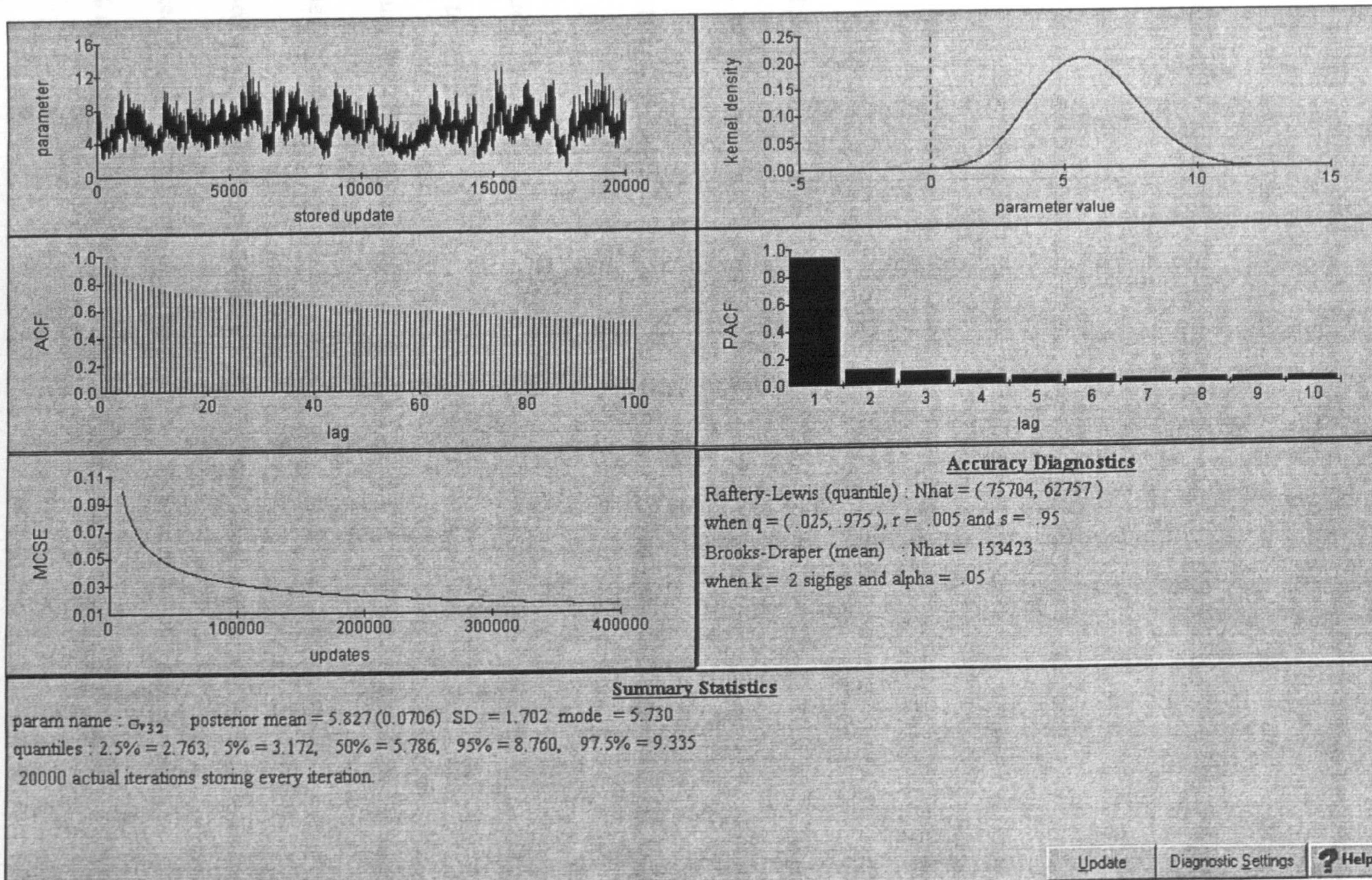
Figure 6.1 MCMC Diagnostics for β_{cms}



Diagnostics for $\sigma_{vallcsw/somecsw}$ are shown in Figure 6.2. The trace for this parameter is 'slower mixing' than the trace for β_{cms} and, as a result, the auto-correlation and partial auto-correlation functions show greater correlations between adjacent values in the chain than were found in Figure 6.1. This lack of independence between successive values in the chain increases the chain length required to satisfactorily estimate this parameter. Raftery-Lewis values of $\hat{N}$ show that a chain length of 75,704 would be required to ensure that the 95% confidence interval for mean β_{cms} would cover between 94% and 96% of the distribution with probability 0.95. However, repeating these calculations after relaxing the accuracy required (with r=0.01) shows that a chain of 18,976 iterations is required to ensure that, with probability 0.95, the 95% confidence interval for $\sigma_{vallcsw/somecsw}$ will cover between 93% and 97% of the distribution and this is viewed as an acceptable level of accuracy.

Diagnostics for the other parameters in the model indicate that a chain length of 20,000 iterations will produce MCMC estimates with an acceptable level of accuracy.

Figure 6.2 MCMC diagnostics for $\sigma_{vallcsw/somecsw}$



6.5 Parameter Estimates

Table 6.2 presents the parameter estimates, for model (6.2), obtained using MCMC estimation. As the sampling distributions of parameters were approximately normal, standard errors can be used here to deduce confidence intervals, although in general this would not be the case. The introduction of the cross-classified part of the model has a number of outcomes. The estimate of the between module variance ($\hat{\sigma}_u^2$ = 9.86, se = 0.7832) confirms that the model based on the cross-classified structure provides a more accurate representation of the data than the simpler hierarchical model used earlier. In the hierarchical models fitted earlier, the variation between modules would have contributed to the estimated variation at level 1. Now that between module variation is represented within the model, a number of level 1 variance terms are reduced: for example, $\hat{\sigma}^2_{econs}$ falls from 57.66 (se = 3.398) in Table 4.10 to 52.07 (se = 2.528) in Table 6.2. Similarly, $\hat{\sigma}_{econs/year1}$, the estimated level 1 covariance term associated with modules taken in the first year falls from 6.571 (se = 0.7934) in Table 4.10 to 4.671 (se = 0.6915) in Table 6.2. There are also changes in the estimated standard errors of some parameters, with the new estimates of standard errors tending to be more conservative than those based on a model which did not recognise the correlations between marks awarded to different students in the same module: for example, the standard error of $\hat{\beta}_{allcsw}$ increases from 0.352 to 0.672. The standard error of another estimate, $\hat{\beta}_{cmsxallcsw}$ increases from 0.695 to 1.284, so that the fixed effect of the interaction between 'computing and mathematical science' and '100% coursework' factors is no longer statistically significant.

Parameter estimates and their associated standard errors are the means and standard deviations of the 20,000 values generated for each parameter by the MCMC estimation process. Percentiles of the posterior distributions can also be derived from the values in the chain. The next section discusses the parameter estimates in more detail. In some cases, functions of one or more parameters are of interest and these have been estimated by calculating the value of the function at each iteration, and then finding the ergodic mean and standard deviation (as described in section 5.5).

Table 6.2 Parameter estimates: MCMC estimation of equation (6.2)

Parameter	Estimate	Standard error
Fixed:		
Constant	54.52	0.845
Project	-2.974	0.607
Male	-1.148	0.4805
International	-2.000	0.6471
Mature	3.485	0.6213
Coffer	-0.0877	0.05764
Professional/managerial	0.2297	0.4811
Year1	0.4526	0.4349
Term3	-0.6228	0.3655
Year3	2.006	0.5312
Stage 2 linear	0.2995	0.1477
Assessment method:		
100%csw	3.027	0.6719
Mixed	2.266	0.6585
Subjects:		
bms	-2.368	0.6722
business	0.9632	0.5248
cms	-0.1212	0.8237
con+es	-1.185	1.043
engineer	-2.006	1.874
h+rm	-0.1049	1.293
human	-0.9224	0.545
lang	-7.422	2.575
planning	1.32	0.7023
law	-1.776	0.9489
Double	0.8979	0.4542
Root class size (centred)	-0.09714	0.07684
Interactions:		
Profman x year3	0.9326	0.3644
Size x year1	0.03378	0.0878
Allcsw x bms	2.046	1.063
Allcsw x cms	-1.312	1.284
Allcsw x engineering	8.262	3.142
Allcsw x h+rm	3.086	1.881
Allcsw x lang	10.56	3.453
Mature x somecsw	-0.8657	0.3241
Somecsw x law	-3.007	1.321
Size x bus	0.1952	0.1083
Size x c+es	-0.3659	0.3156
Size x h+rm	0.7349	0.333
Size x law	-0.2265	0.1965
male x year1	0.6623	0.3592
mature x st2linear	-0.3483	0.1107
takeout	-34.13	0.7208

Random:		
Student level: σ^2_{vcons}	36.65	4.304
$\sigma_{vcons/allcsw}$	-15.01	2.984
$\sigma^2_{vallcsw}$	13.2	2.636
$\sigma_{vcons/somecsw}$	-6.731	2.145
$\sigma_{vallcsw/somecsw}$	5.827	1.702
$\sigma^2_{vsomecsw}$	3.087	1.077
$\sigma_{vcons/year1}$	0.9308	1.361
$\sigma_{vcyear1/allcsw}$	-3.374	1.115
$\sigma_{vyear1/somecsw}$	-2.505	0.8484
σ^2_{vyear1}	5.922	0.966
σ^2_{vyear3}	7.322	0.9716
$\sigma^2_{vmature}$	14.54	4.863
Between modules σ^2_u	9.856	0.7832
Level 1 variance		
$\sigma^2_{[illegible]}$	52.07	2.528
Covariance between level 1 residuals associated with cons and:		
project	-0.7334	1.132
100% coursework	-12.06	1.297
mixed assessment	-14.15	1.298
bms	16.11	1.365
bus	9.653	0.9519
cms	12.22	2.755
con+es	6.662	1.349
engineer	29.86	9.432
lang	25.37	8.405
workload 5+	5.015	1.543
male	4.01	0.5604
mature	-1.602	0.5677
class size	-0.3664	0.1494
year1	4.671	0.6915
stage2 basic	12.38	2.984
size xyear1	-0.07401	0.212
bms x allcsw	-10.81	2.939
bus x allcsw	-7.76	1.521
cms x somecsw	32.25	4.496
law x somecsw	4.338	1.198

6.6 Effects of Student Characteristics

6.6.1 Sex

During stage 2 (the second and third years), the results achieved by male students were on average 1.15 marks (se = 0.481) lower than those achieved by female students. In an earlier analysis, the difference between male and female students' performance was significantly smaller in the first year, but after the introduction of the variation between modules, this effect is no longer statistically significant ($\hat{\beta}_{maleyr1}$ =0.66, se = 0.359).

Within programmes, male students' marks were more variable than female students'. The level 1 variance is $\sigma^2_{econs} + 2\sigma_{econs/male}$ for male students and σ^2_{econs} for female students: so the estimated variance for male students is $2\hat{\sigma}_{econs/male}$ = 8.02 (se = 1.120) units higher than the variance for female students.

These results show that, in general, male students performed at a lower level and less consistently than similar female students in the same types of modules. The lower mean marks achieved by male students will lead to lower averages in the calculation of degree class, although this will be alleviated to some extent by the effect of greater variation at level 1. As explained in section 2.6, a student whose performance varies could gain more from the selection of their best 16 module marks to determine degree class and from the operation of both routine and borderline classification procedures than a more consistent student with the same mean level of performance overall. Another consequence of the lower level of achievement and greater variation at level 1 is that men are more likely than women to fail modules, which could mean having to take heavier workloads in order to compensate, and a greater chance of being awarded an ordinary degree. Overall, this suggests that, other things being equal, male students' degree classifications will be more varied than female students'.

It is worth noting that among the effects which did not appear to depend on gender are those of assessment method, subject, progress and age, as neither the current analysis nor earlier analyses found no evidence of interactions between gender and these factors.

6.6.2 Domicile

International students had lower mean performance than home students of the same age and background under the same module conditions: international students' marks were on average 2.00 (se = 0.647) marks lower than those of home students of the same age, gender and social class in similar modules

6.6.3 Entry Qualifications

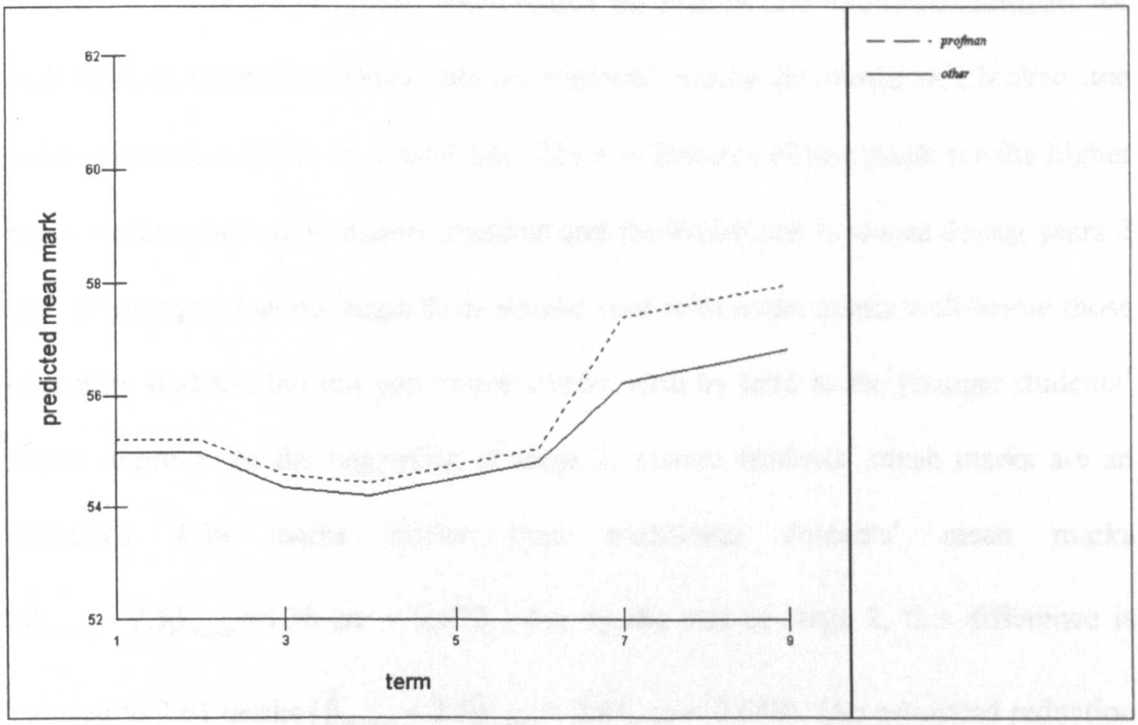
Recall that, for reasons explained in section 4.5.1, the measure of entry qualifications used in the analyses (COFFER), reflects the university's expectations of the student rather than the student's own achievements before entering the Modular Degree Programme. This measure had no significant effect on students' mean marks $\left(\hat{\beta}_{coffer} = -0.09, se = 0.058\right)$. This could be a consequence of using a measure of minimum expectations rather than actual prior attainment, and of ascribing the same value for this measure to all students studying for the same degree, but other studies which have measured students' entry qualifications directly have found only weak links between A-level results and subsequent performance in higher education as discussed in section 2.3.3.

6.6.4 Social Class

Although initially (Table 4.4), it seemed that there was no social class effect on performance, the introduction of interaction terms (see Table 4.5) identified a difference between social classes in the 'step up' in achievement occurring between the second and third years and which is also evident in Table 6.2. Students whose family included a parent with a professional or managerial occupation had similar performance on average to other students during the first and second years ($\hat{\beta}_{profman}$=0.23, se = 0.481), however, the 'step-up' between second and third years was 0.93 marks higher (se = 0.364) for students with a parent in a professional or managerial occupation. The size of this estimated difference is sufficient, other things being equal, to produce some differences in degree classifications. Students with a professional or managerial background will have higher average marks in modules contributing to degree class and will be more likely to meet the criteria defining 'borderline' students eligible for upgrading to a higher degree class, which requires that a majority of the last modules counting toward degree class are in the range corresponding to the higher class. The greater a student's 'step up' between years 2 and 3, the more likely this condition is to be met.

In Figure 6.3 shown below, predicted marks for students with a parent in a professional/managerial occupation are plotted as a dashed line, and predicted marks for other students shown as an unbroken line. In the first and second years, there is no statistically significant difference between classes ($\hat{\beta}_{profman} = 0.23, se = 0.481$), but in the third year, mean marks are significantly higher for students with a parent in a professional or managerial occupation, because of their greater 'step up' in achievement $\left(\hat{\beta}_{pmyear3} = 0.9326, se = 0.3644\right)$.

Figure 6.3 Estimated mean marks, by term, according to parental occupation



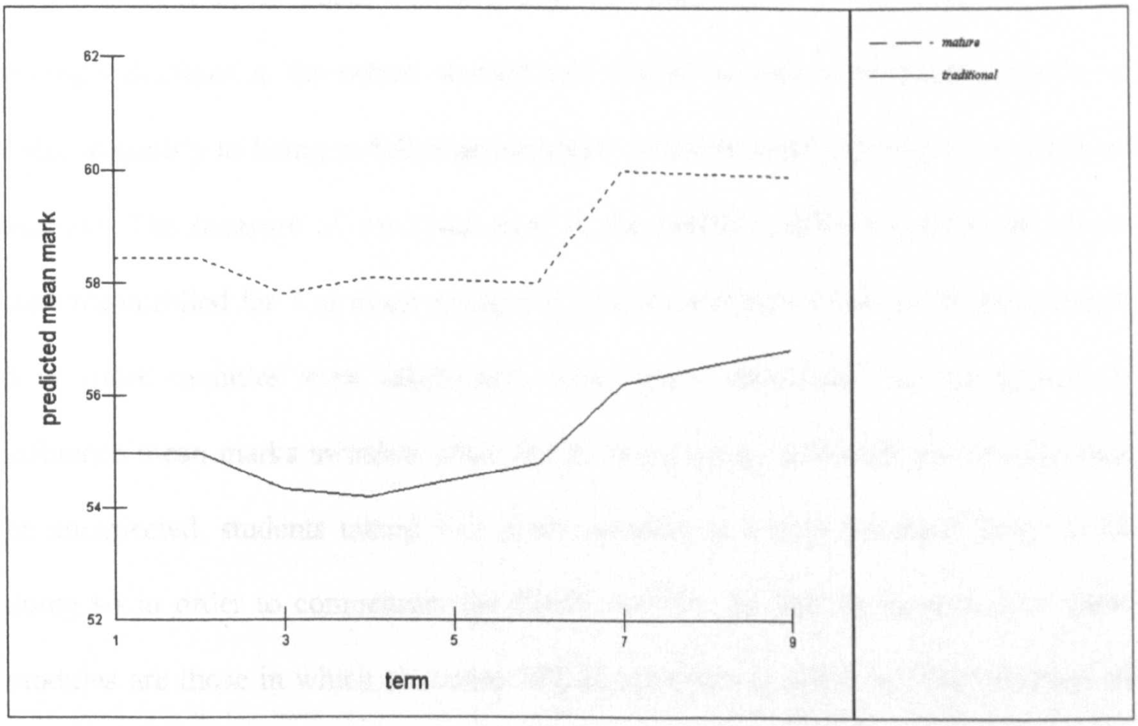
6.6.5 Age

The analysis shows that the relationship between age and achievement depends on the method of assessment used and on the point within a student's programme at which comparisons are made. Mean marks achieved by mature students in modules assessed by either 100% coursework or 100% examination are 3.49 (se = 0.621) marks higher than those achieved by students who entered the Modular Degree Programme at the traditional age . This difference is reduced by 0.87 (se = 0.324) marks when the assessment is a mixture of coursework and examination, but even in these modules mature students still achieve considerably higher average marks than traditional students ($\hat{\beta}_{mature} + \hat{\beta}_{matsome} = 2.62$, se = 0.649).

There is also an interaction between student's age and the linear element of students' progress: $\hat{\beta}_{matlin}$=-0.348 (se = 0.111), showing that, on average, mature students make less progress term by term during stage 2, than traditional students. Figure 6.4 shows the estimated mean marks for mature and traditional students for each term, in standard modules. Mature students' means are shown as a broken line and traditional students' as a solid line. The key features of this graph are the higher mean marks achieved by mature students and the difference in slopes during years 2 and 3: younger students begin their second year with mean marks well below those of mature students, but this gap closes a little, term by term as the younger students' marks improve. At the beginning of stage 2, mature students' mean marks are an estimated 4.36 marks higher than traditional students' mean marks ($\hat{\beta}_{mature} - 2.5\hat{\beta}_{matlin}$ =4.36 ,se = 0.672) but by the end of stage 2, this difference is reduced to 2.61 marks ($\hat{\beta}_{mature} + 2.5\hat{\beta}_{matlin} = 2.61$, se = 0.688). (An estimated reduction of $-5\hat{\beta}_{matlin} = 1.7$ marks, se =0.555). Hence, although the gap narrows, in all nine terms, the mean for younger students is significantly below the mean for mature students, other things being equal.

A further difference between mature and traditional students is in the level 1 variation, $2\hat{\sigma}_{emature}$ = -3.20 (se = 1.135) units, so that the level 1 variation is lower for mature students. If mature students perform more consistently, this suggests that they may be better able to deal with differences in style/resources between modules than younger students. With a higher mean and greater consistency, mature students appear to be more in control of their performance are less likely than younger students to fail modules.

Figure 6.4 Estimated mean marks, by term, for mature and traditional students



So far, comparisons have been made between 'average' students in different age groups; complex variation at level 3 showed that there was greater variation between mature students than between students entering the course at the traditional age. The between student variance is $\hat{\sigma}^2_{vmature}$ = 14.54 units (se = 4.863) higher for mature students. A more detailed discussion of this finding will be given in section 6.8.

6.6.5 Workload

The number of modules taken each term depends on how a student chooses their programme of study. Initially determined by a student's choices within the restrictions governing their particular degree, a student may change their programme as their preferences develop or in response to the results achieved in earlier modules. Some students may add modules to their programme in order to compensate for earlier failures or in hopes of being able to 'drop' poor marks from their final average. Students in the cohort studied here needed to take 3 modules per term in order to qualify as being in full-time education and 4 modules per term would not be unusual. The measure of workload used in the analysis identified terms in which students enrolled for 5 or more modules. Comparing marks achieved in terms where 5 or more modules were taken with other terms, workload did not appear to influence mean marks awarded, other things being equal. Although this finding may be unexpected, students taking 5 or more modules in a term are most likely to be doing so in order to compensate for failed modules, so that terms with 5 or more modules are those in which strenuous efforts are made to catch up. The selection of the sample means that only students who succeeded in making up for past failures and graduating without having to extend beyond three academic years are represented. This means that the analysis provides only a limited opportunity to examine the effects of workload on performance, and being based on cases where students were successful in rescuing their programme, will tend to provide an optimistic view of the effects of workload.

6.7 Effects Of Module Characteristics

One objective of the analysis was to determine the extent to which students' performances vary because the programmes selected by different students lead them to take different types of modules. Module characteristics whose effects on performance were studied were: method of assessment (100% coursework/100% examination/some coursework), module size (double/single), subject group, class size, type (project/other), term, year, level and the interactions between class size and subject, assessment method and subject. These characteristics were represented in the model as potentially generating fixed effects, effects varying between students and/ or defining complex variation at level 1. The between module variation, σ^2_u, measures any unexplained variation between the marks awarded in different modules.

6.7.1 Assessment Methods

The choice of assessment method was found to impact on students' marks in a variety of ways. On average, higher marks were achieved in modules using a mixture of assessment methods and in modules using 100% coursework assessment than in modules using only examination assessment ($\hat{\beta}_{allcsw}$=3.03, se = 0.672: $\hat{\beta}_{somecsw}$=2.27, se = 0.659). Compared to these effects, the difference between the mean marks achieved in 100% coursework and mixed assessment is relatively small (0.76, se = 0.41).

The impact of mixed coursework and examination assessment, compared to 100% examination assessment was greater for traditional students, who had entered the Modular Programme aged 20 or below. For these students, the estimated mean performance in modules using a mixture of assessment methods was 2.27 marks (se

= 0.659) higher than in modules using 100% examination, whereas for mature students the estimated mean difference is 1.4 marks (se = 0.695). As a result, for mature students, the mean marks achieved in 100% coursework modules are 1.63 marks higher (se = 0.47) than the mean marks achieved in modules using mixed assessment.

The impact of different forms of assessment was also dependent on the subject group administering the module; these effects will be described in the next section.

Assessment methods contributed to complex variation at level 1: the estimated level 1 variances for modules using different forms of assessment are shown in Table 6.3. Use of coursework assessment is associated with large reductions in the level 1 variation, other things being equal.

Table 6.3 Level 1 variances in modules using different forms of assessment

	Estimated level 1 variance	Standard error	95% confidence interval
100% examination	52.1	2.53	(47.1, 56.9)
Some coursework	23.8	1.04	(21.7, 25.8)
100% coursework	28.0	1.16	(25.7, 30.3)

Section 6.8 will describe how individual students varied in their responses to different assessment methods.

6.7.2 Subjects

As expected, there were significant differences between subjects in terms of mean marks achieved and the level 1 variation, reflecting the differences in mark distributions known to exist between different disciplines and discussed in chapter 2. Differences in the mark distributions corresponding to different disciplines are incorporated within the model as fixed effects and complex variance terms at level 1 and interactions between subjects and assessment methods had statistically significant effects on both mean marks and level 1 variances.

As assessment practices varied between subjects, and there were interactions between subject and assessment methods, comparisons between subjects are made separately for each type of assessment regime. The distribution of module entries by assessment method by subject shows that in the modules taken by this sample, not all of the subject groups used all three assessment options: categories in which there were fewer than 50 entries are excluded from the reports below. Figures 6.5-6.7 plot the predicted mean marks achieved in 'standard' modules by 'standard students' by subject, for each form of assessment, with 95% confidence intervals calculated separately for each mean.

Figure 6.5 Predicted mean marks, by subject, for 'standard students' in 'standard' modules using 100% coursework assessment

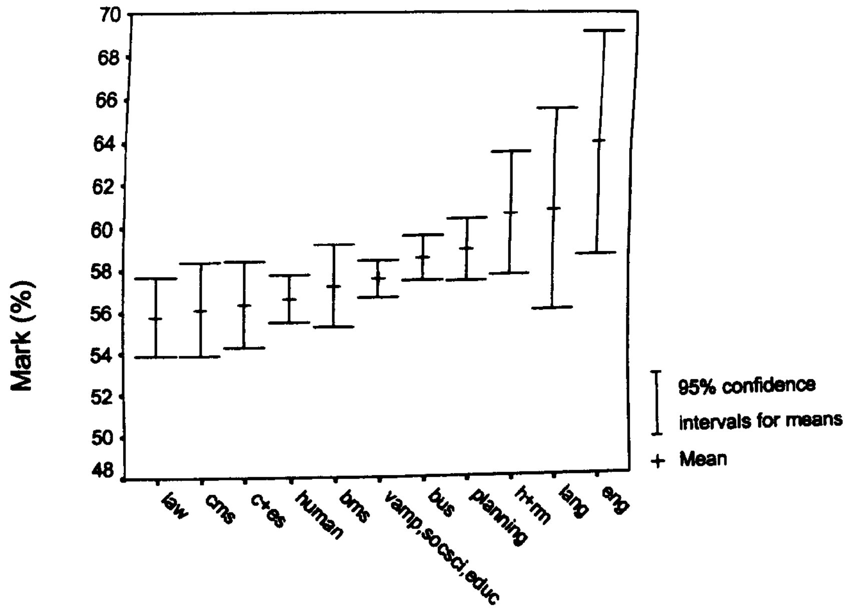


Figure 6.6 Predicted mean marks, by subject, for 'standard students' in 'standard' modules using mixed coursework and examination assessment

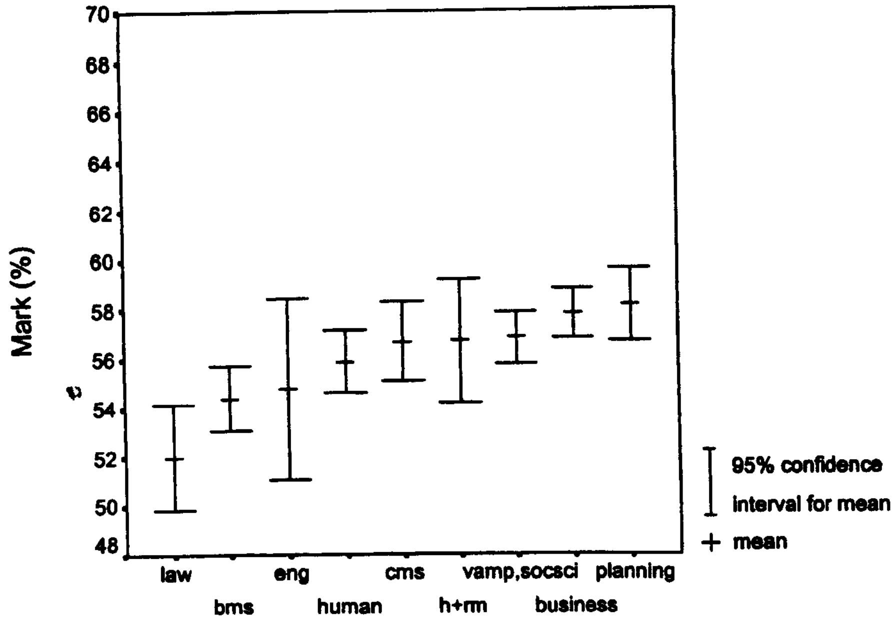


Figure 6.7 Predicted mean marks, by subject, for 'standard' students in 'standard' modules using 100% examination assessment

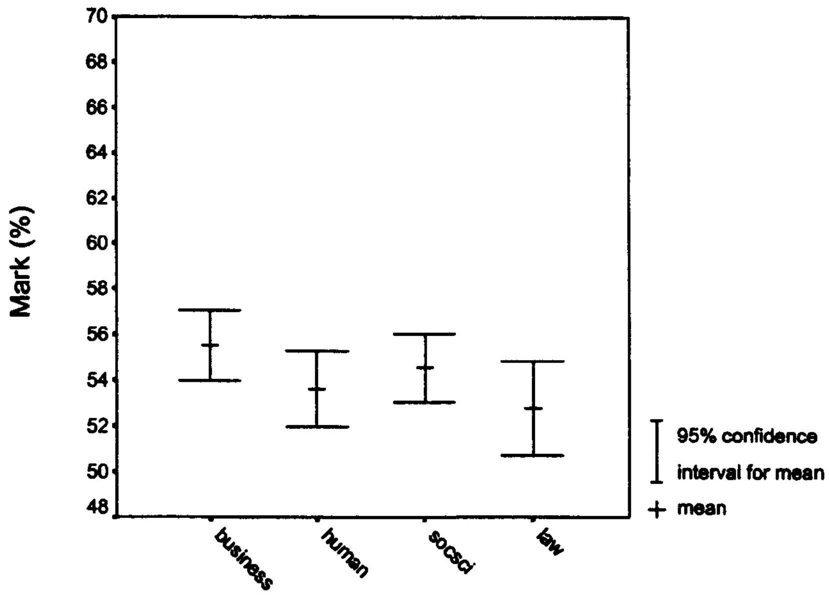


Table 6.4 shows the difference between the mean marks achieved in modules using 100% coursework assessment compared to means achieved in modules using a mixture of coursework and examination assessment, by subject, other things being equal. This table shows that there were variations between subjects in the mean differences between the marks awarded on the basis of 100% coursework assessment those based on mixed assessment. In Biology and Molecular Sciences, Hotel and Restaurant Management and Law, mean marks achieved in 100% coursework modules were significantly higher than in modules using mixed assessment. In other subjects, no significant increase in mean marks was associated with assessment by coursework rather than a mixture of assessment methods.

Figure 6.8 plots level 1 variances by fixed effects by subject for 100% coursework assessment. Notice that in Computing and Mathematical Sciences, modules have high level 1 variation in 100% coursework modules and still higher level 1 variation in modules using a mixed assessment regime. Figure 6.10 shows that the marks awarded in Business modules using 100% examination assessment have higher level 1 variation than the marks awarded in other subjects using this form of assessment. Level 1 variation in Business modules is also higher than in these other subjects when assessment is carried out using a mixture of methods.

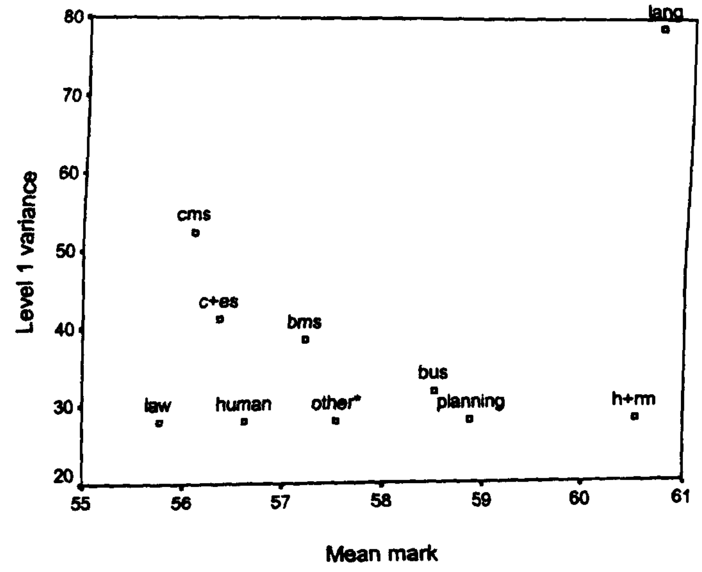
As no account was taken of the relative weights given to the coursework and examination element in modules using a mixture of assessments, it is possible that some of the interactions between subjects and assessment methods are due to the allocation of different weights to each element in modules using both forms of assessment. Interactions could also reflect subject differences in the nature of the coursework assignments: for example, programming in Computing modules, experimental or laboratory work in Biology and Molecular Sciences, essays in Social Sciences, case studies in Hotel and Restaurant Management.

Table 6.4 Estimated mean marks awarded in modules using 100% coursework minus estimated mean marks awarded in modules using mixed assessment, by subject, standard modules and students

Subject	Mean difference	se	95% confidence intervals for difference
Bms	2.806	1.015	(0.784, 4.786)
Cms	-0.552	1.238	(-2.950, 1.892)
H+rm	3.846	1.843	(0.266, 7.475)
Law	3.767	1.312	(1.181, 6.355)
Other*	0.76	0.410	(-0.043, 1.566)

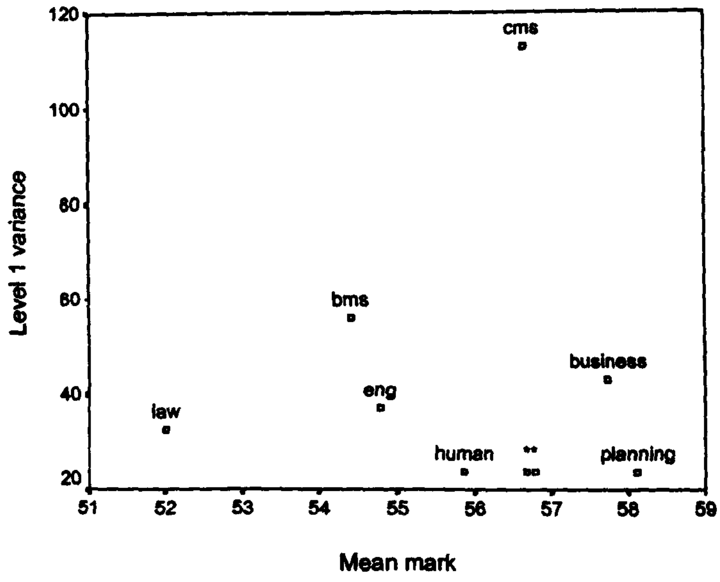
*social science, vamp, educ, business, humanities, con+es, planning

Figure 6.8 Relationship between estimated mean and level 1 variance by subject, modules using 100% coursework assessment



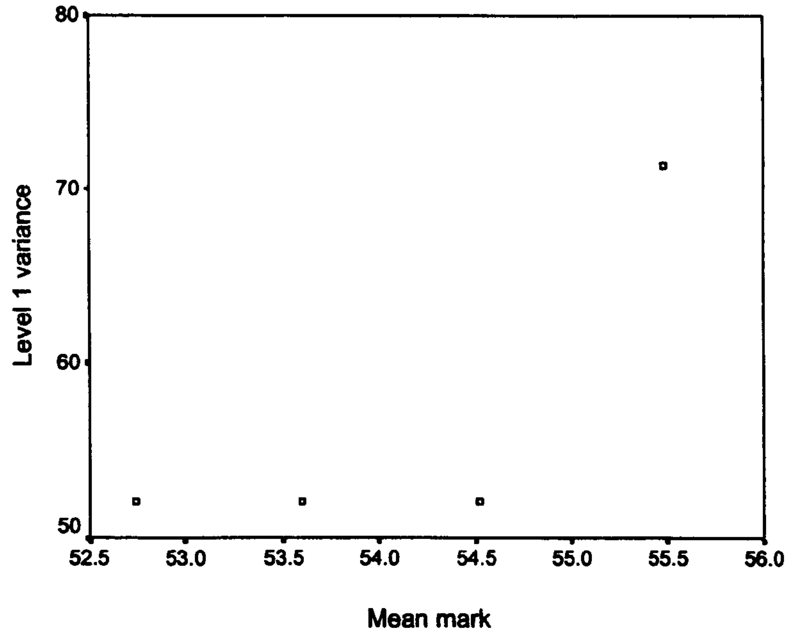
*vamp, social science, education

Figure 6.9 Relationship between estimated mean and level 1 variance, by subject, in modules using a mixture of coursework and examination assessment



*vamp,socsci, h+rm

Figure 6.10 Relationship between estimated mean marks and level 1 variance by subject, modules using 100% examination assessment



6.7.3 Projects

Final year projects or dissertations are double modules which students must pass in order to graduate with honours. Assessment is based on a written report of up to 10,000 words, and carried out by two assessors who must agree a mark. Where students have carried out an interdisciplinary project, one person from each field will assess their work. Other things being equal, project marks are on average of 2.97 marks lower (se = 0.607) than marks achieved in other modules with the same characteristics. Note that projects are 100% coursework modules taken in the final year, and are therefore compared to modules in which students tend to achieve their highest marks. Although included amongst the variables defining complex variation at level 1, after allowing for variation between modules, the estimate of complex variance term $(\sigma_{econs/projects})$ associated with projects is small relative to its standard error (-0.73,se = 1.13). Although many students will have experience of 100% coursework assessment in other modules, these modules will have made different demands to those made by an extended piece of independent work, largely defined by the student, such as the final year project.

6.7.4 Single and Double Modules

The difference between the mean marks awarded in double modules compared to those awarded in single unit modules $(\hat{\beta}_{double} = 0.90, se = 0.454)$ was not statistically significant.

6.7.5 Class Size

The mean marks predicted by the model for a standard student in a standard module with N students enrolled is $\beta_{cons} + \beta_{crtn}\left(\sqrt{N} - 7.837\right)$. Other things being equal, class size (or module enrolment) was not significantly related to mean marks $\left(\hat{\beta}_{crtn} = -0.097, se = 0.0768\right)$, but there was some evidence of a different relationship between class size and mean marks for modules administered by hotel and restaurant management. In these modules predicted marks are $\beta_{cons} + \beta_{h+rm} + (\beta_{crtn} + \beta_{crtsizexh+rm})\left(\sqrt{N} - 7.837\right)$ and $\beta_{crtsizexh+rm}$ has an estimated value of 0.737 (se = 0.333). Figures 6.11 and 6.12 plot predicted mean marks according to module enrolment, for all subjects and for hotel and restaurant management respectively, with 95% confidence limits for the means. (Note the limited range of enrolment numbers in Figure 6.12).

Figure 6.11 Predicted mean marks, with 95% confidence intervals, by module enrolment: standard modules in stage 2

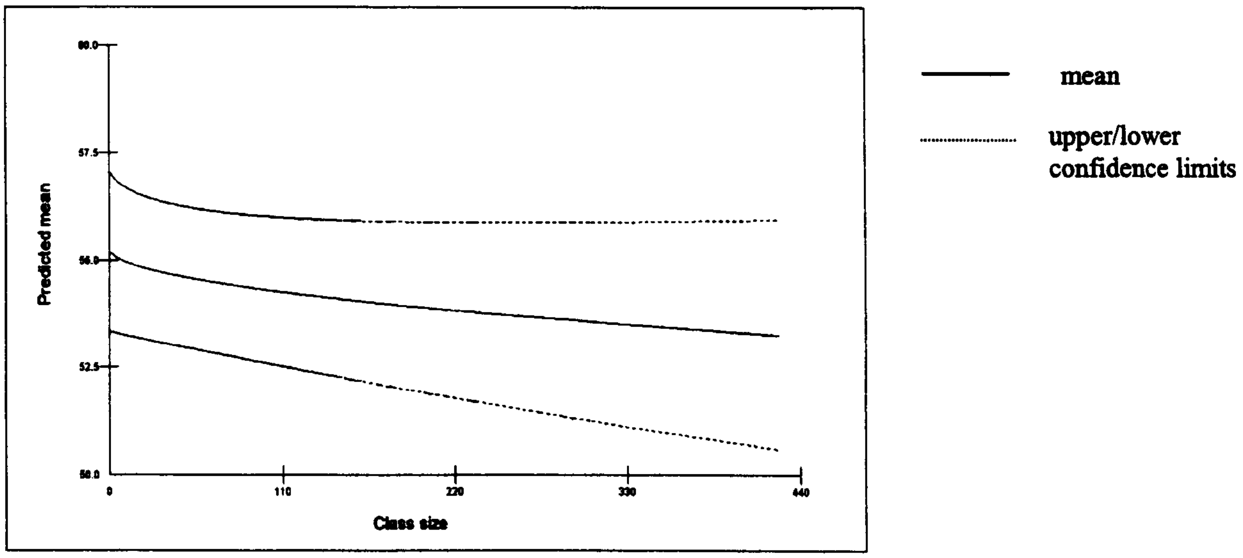
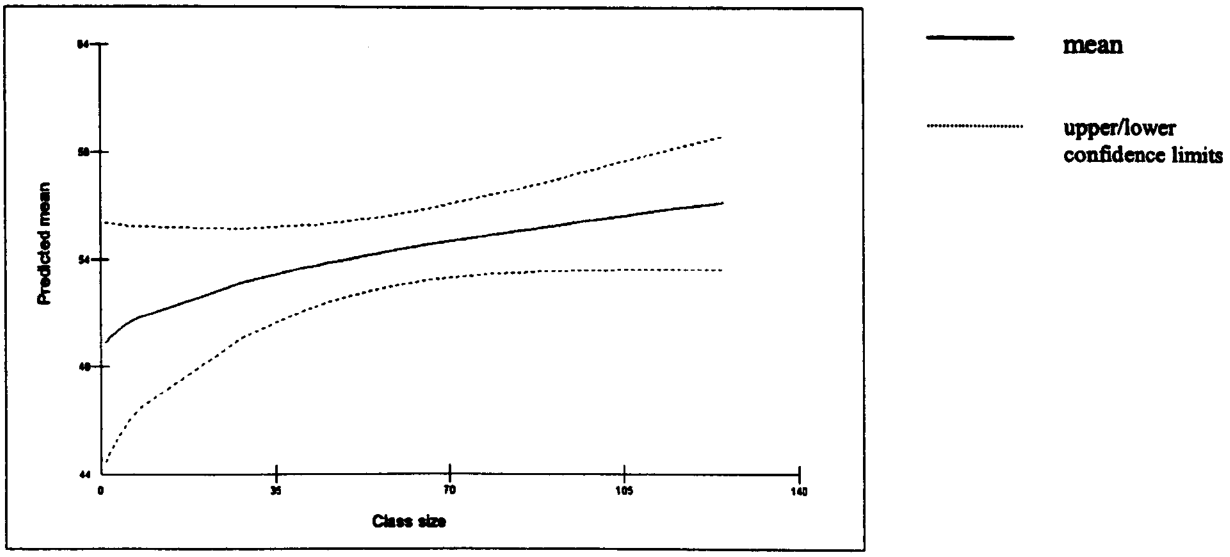


Figure 6.12 Predicted mean marks, with 95% confidence intervals, by module enrolment: standard modules in hotel and restaurant management



6.7.6 Term and Level

Section 4.5.1 described how the particular form for modelling mean student progress was selected. After dropping terms that did not make a significant contribution to earlier models, four parameters were used to represent mean progress. For standard students, taking standard modules, the mean marks achieved in term t are given by the following expression:

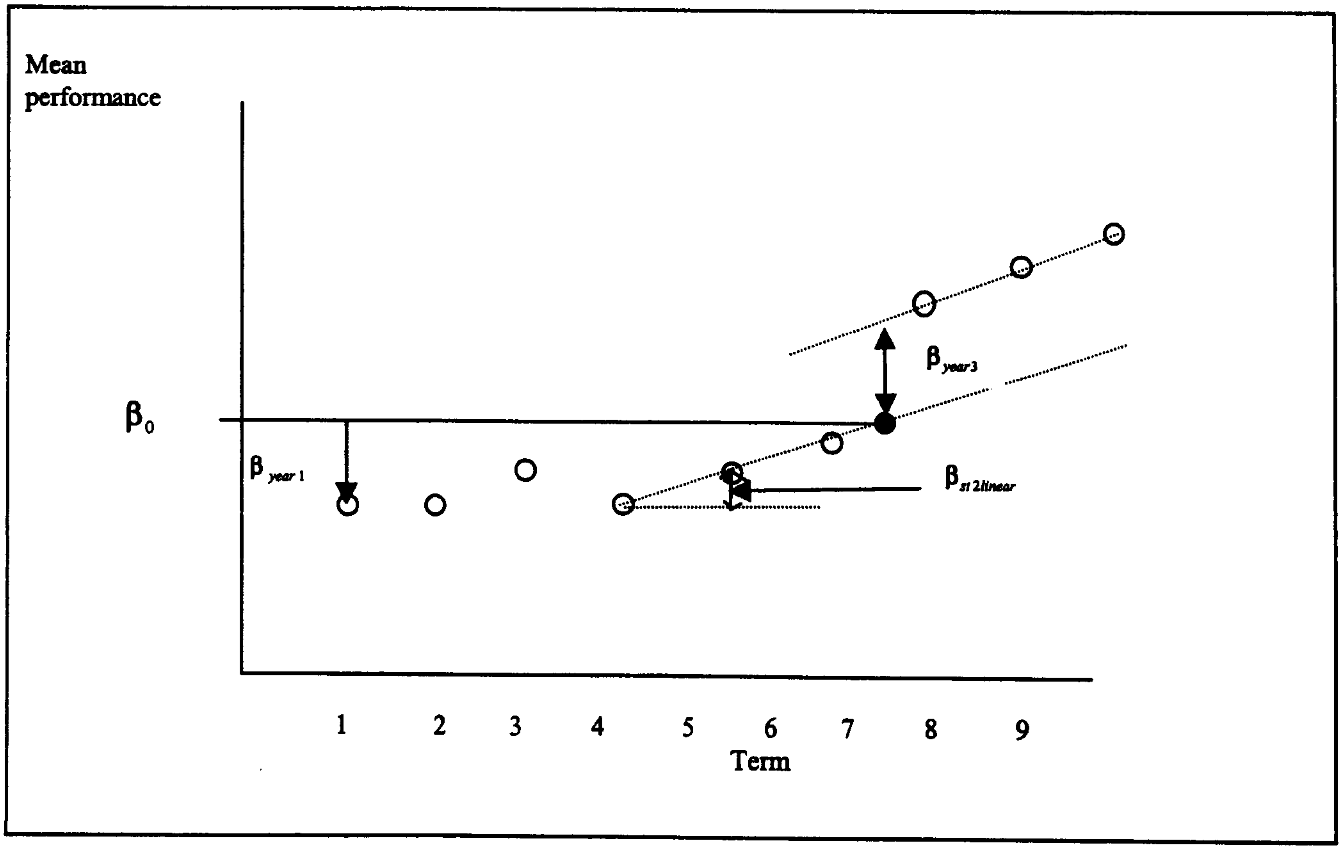
$$\beta_{cons} + \beta_{year1}\text{YEAR1} + \beta_{term3}\text{TERM3} + \beta_{year3}\text{YEAR3} + \beta_{st2linear}(t-6.5)*(1-\text{YEAR1})$$

where YEAR1, YEAR3 and TERM3 are dummy variables coded 1 for the periods indicated by the variable name and 0 otherwise.

This expression is shown below, evaluated for each term and is plotted in Figure 6.13. Results given in Table 6.2 show an average student's progress can be described as improving linearly during stage 2, with a 'step up' between the second and third years.

Term:	**Mean marks: standard student, standard module**
1	$\beta_{cons} + \beta_{year1}$
2	$\beta_{cons} + \beta_{year1}$
3	$\beta_{cons} + \beta_{year1} + \beta_{term3}$
4	$\beta_{cons} - 2.5\beta_{st2linear}$
5	$\beta_{cons} - 1.5\beta_{st2linear}$
6	$\beta_{cons} - 0.5\beta_{st2linear}$
7	$\beta_{cons} + 0.5\beta_{st2linear} + \beta_{year3}$
8	$\beta_{cons} + 1.5\beta_{st2linear} + \beta_{year3}$
9	$\beta_{cons} + 2.5\beta_{st2linear} + \beta_{year3}$

Figure 6.13 Parameters describing student progress



Two parameters are related to changes in mean performance between terms: β_{term3} measures the change between the first two terms and final term of the first year. This parameter was not statistically significant (-0.62, se = 0.366). Another parameter, β_{st2lin}, measures linear progress during stage 2: the estimate of β_{st2lin} is 0.30 marks (se = 0.148). This shows that on average, students' mean marks improve as they become more experienced.

Two parameters are related to changes in mean performance between academic years: β_{year1} compares mean performance in terms 1 and 2 of the first year with the mean during hypothetical 'standard' term in year 2. The change in mean marks between the last term of the first year and the first term of the second year

$(-\beta_{term3} - \beta_{year1} - 2.5\beta_{st2linear})$ shows the immediate effect of moving from basic modules taken in the first year to mainly advanced modules taken in stage 2 more directly. This has an estimated value of -0.579 (se = 0.457), but, as section 6.8 will show, varied between students.

β_{year3}, with an estimated value of 2.01 (se = 0.531) represents the mean improvement between the end of the second year and the beginning of the third, over and above usual term by term progress. The estimated value of this parameter $\left(\hat{\beta}_{year3} = 2.006, se = 0.5312\right)$ indicates a substantial 'step up' in students mean achievement between the second and third years. This improvement could be the result of students consolidating the knowledge gained in the second year or could show the effects of greater motivation in the final year as students focus on the impact of their performance on their final classification. Allowing this parameter to vary between students showed that there was significant variation between students in the progress made between the second and third years. This will be discussed in section 6.8.

Patterns of progress were different for some sub-groups of students: those from a professional or managerial background had a higher step up in mean marks between years 2 and 3 and those entering at the traditional age made greater progress from term to term during stage 2 than mature students. These variations have been described in section 6.6.4 and 6.6.5. There were also significant variations between individual students in the progress made from one academic year to another and these variations will be described in section 6.8.

Estimates of variance parameters showed that the level 1 variation also changed during the course of students' programmes. For 'standard' students in 'standard' modules, level 1 variation is $\left(\sigma^2_{econs} + 2\sigma_{eyear1}\right)$ in the first year and σ^2_{econs} in the second and third years. The estimate of the covariance term, σ_{eyear1}, was 4.67 (se = 0.692) hence for 'standard' modules, taken in the first year, the estimated level 1 variation is 61.41 (se = 2.657).

A number of reasons may explain why students' performances in the first year are more erratic than later on in their programmes. During the first year, students have less experience in balancing their efforts across several modules and although they need to meet conditions for progressing to stage 2, the marks they achieve do not otherwise count towards the classification of final award. First year students are therefore potentially less able and less motivated to perform consistently in all their modules. It is also possible that the greater diversity of first year programmes compared to later ones contributes to the within student variability in marks.

The progression from stage 1 to stage 2 is accompanied by a switch from basic to advanced modules, so that in general the effect of 'level' is confounded with time. The exception to this is that the regulations allowed these students to include up to 2 basic modules during their stage 2 programmes. The difference between mean performance in 'basic' and 'advanced' modules taken in stage 2 was measured by the parameter $\beta_{stage2basic}$. This parameter, which was not significantly different from zero, was eliminated at an earlier stage in building the model, but the effect of 'level' remained as contributing to complex variation at level 1, with $\sigma_{econs/st2basic}$ estimated to be 12.38 (se = 2.984). This means that although there was no difference between basic and advanced modules taken during stage 2 in the mean level of marks awarded, the marks awarded in basic modules were more varied. This could reflect different strategic approaches taken by students putting basic modules into their stage 2 programmes: one strategy being to use basic modules to rescue failing records in the hope that adding 'easy to pass' basic modules would help to gain module credits necessary for graduation and another being to use basic modules as an opportunity to achieve high marks in order to improve on predicted degree classification.

6.8 Student Level Random Effects

The purpose of including simple level 3 variation within the model is to recognise that, after allowing for differences in background and the types of modules taken, students will have different mean levels of performance. Five coefficients of predictor variables (other than the intercept) were allowed to have random effects at level 3 (students). The introduction of residuals associated with these variables allows the effects of assessment methods and time to vary from student to student and introduces a complex variance term depending on student's age. These effects are measured by the variances and covariances that make up the matrix Ω_v and were identified earlier in section 6.3.

According to model (6.2), the mean performance for student *k*, in 'standard' modules, after adjusting for the explanatory variables, is $\beta_{cons} + v_{consk}$; this has mean β_{cons} and variance σ^2_{vcons}. The estimate of σ^2_{vcons} shows that in standard modules, 'standard' students' means had estimated variance 36.65 (se = 4.304). The variation between student means was larger amongst mature students by an amount $\hat{\sigma}^2_{mature}$ = 14.5 (se = 4.86). Table 6.5 shows the variation in students' means for standard modules, by age.

Table 6.5 Estimated variance of student means: traditional and mature students

	Estimate	Standard error	95% confidence interval
Traditional students	36.7	4.30	(28.7, 45.6)
Mature students	51.2	5.80	(40.9, 63.8)

Other random effects were introduced at student level in order to individualise the effects of different assessment methods and patterns of progress over time. In model (6.2), the effect of 100% coursework assessment, contrasted with the 100% examination assessment, on the mean marks of student k is $\beta_{allcsw} + v_{allcswk}$. These effects are normally distributed with mean β_{allcsw}, (estimate. 3.027, se = 0.672) and variance $\sigma^2_{vallcsw}$ (estimate = 13.2, se = 2.636). The proportion of students for whom the mean increase in marks associated with coursework assessment lies in a given range is a function of the model parameters, β_{allcsw} and $v_{allcswk}$. Using equation (5.9) on page 181, any function of the model parameters can be estimated from the chain of sampled parameter values, hence the value of $\Phi\left(\dfrac{x-\beta_{allcsw}}{\sigma_{vallcsw}}\right)$, where Φ is the area under a standard normal curve, can be estimated for any value of x. Using this approach we find that an estimated 61% of students (95% confidence interval: 47%-75%) would be expected to gain 2 marks or more on average in modules using coursework rather than examination assessment. Not all students gain from coursework assessment: 20% of students (95% confidence interval 10%-32%) would be expected to perform better, on average, in modules using examination assessment.

For modules using a mixture of coursework and examination, but otherwise 'standard', a 'standard' student k will achieve a mean mark that is $\beta_{somecsw} + v_{somecswk}$ marks higher than in modules based wholly on examination assessment. The estimated mean of this difference is 2.27 (se = 0.659) and the size of this effect varies from student to student with variance $\sigma^2_{vsomecsw}$ (estimated value =3.09, se = 1.077) . With these results, an estimated 56% (95% confidence interval : 26%-83%) of students would be expected to gain at least two marks from mixed assessment rather than assessment by examination and 10% of students (95% confidence interval: 1% - 30%) would be expected to perform better in modules relying on examination assessment with no coursework component.

For a 'standard' student k, the mean marks achieved in 100% coursework modules will be $(\beta_{allcsw} + v_{allcswk} - \beta_{somecsw} - v_{somecswk})$ higher than their mean in modules using mixed assessment.

This difference has an estimated mean of 0.76 (se = 0.410) and a between student variance of $(\sigma^2_{vallcsw} + \sigma^2_{vsomecsw} - 2\sigma_{vallcsw/somecsw})$. This has an estimated value of 4.63 (se = 0.881), showing that the gains from avoiding an examination tend to be smaller when the examination is only part of the assessment for a module. 28% of students (95% confidence interval: 16% - 42%) will gain an average of 2 or more marks from the removal of the examination component of the assessment in modules using a mixture of coursework and examination.

Table 6.6 shows the estimated variance between students in modules using different assessment methods but which are otherwise similar. The variation in the responses of individual students to different forms of assessment mean that the variation between students in mean performance depends on the form of assessment used: coursework assessment tends to reduce the variation between students.

When coursework is added to or replaces examination assessment, average marks awarded increase, the consistency of student performances improves and variation between students falls. These effects would be expected to lead to substantially different mark distributions for students whose programmes differed in terms of assessment regimes associated with their chosen modules, but who were otherwise similar.

Table 6.6 Variation between students in mean performance by assessment method, standard modules taken in stage 2

Method of Assessment	Estimate	Standard error	95% confidence interval
100% examination $\left(\sigma^2_{vcons}\right)$	36.7	4.30	(28.7, 45.6)
Some coursework $\left(\sigma^2_{vcons} + \sigma^2_{vsomecsw} + 2\sigma_{vcons/somecsw}\right)$	26.3	2.21	(22.2, 30.9)
100% coursework $\left(\sigma^2_{vcons} + \sigma^2_{vallcsw} + 2\sigma_{vcons/allcsw}\right)$	19.8	1.91	(16.3, 23.8)

In the first year the between-student variation in the marks achieved in 'standard' modules is $\left(\sigma^2_{vcons} + \sigma^2_{vyear1} + 2\sigma_{vcons/year1}\right)$, compared to $\left(\sigma^2_{vcons}\right)$ in the second year. The variation between students in stage 2 has an estimated value of 36.65 (se = 4.30); the fall in the variation between students is statistically significant and has been adjusted for the effects of factors such as assessment method and class size, which might be expected to produce this kind of effect ($2\hat{\sigma}_{vyear1/cons} + \hat{\sigma}^2_{vyear1} = 7.78, se = 2.767$, 95% confidence interval 2.42, 13.37).

The residuals $\{v_{consk}\}$, $\{v_{allcswk}\}$and $\{v_{somecswk}\}$are related to each other, with correlation coefficients given by the formulae below. These correlation coefficients were estimated by calculating the values for each of the functions below for each iteration in the chain. The values obtained were then used to calculate the estimates, standard errors and 95% confidence intervals for the correlation coefficients and these are presented in Table 6.7.

$$\rho(v_{cons}, v_{allcsw}) = \frac{\sigma_{vcons/allcsw}}{\sqrt{\sigma^2_{vcons}\sigma^2_{vallcsw}}}$$

$$\rho(v_{cons}, v_{somecsw}) = \frac{\sigma_{vcons/somecsw}}{\sqrt{\sigma^2_{vcons}\sigma^2_{vsomecsw}}}$$

$$\rho(v_{somecsw}, v_{allcsw}) = \frac{\sigma_{vsomecsw/allcsw}}{\sqrt{\sigma^2_{vsomecsw}\sigma^2_{vallcsw}}}$$

Table 6.7 Correlations between individual student's responses to different forms of assessment, standard students and modules

Correlation coefficient	Estimate	Standard error	95% confidence interval
$\rho(v_{cons}, v_{allcsw})$	-0.68	0.055	(-0.78,-0.56)
$\rho(v_{cons}, v_{somecsw})$	-0.631	0.108	(-0.81, -0.38)
$\rho(v_{somecsw}, v_{allcsw})$	0.912	0.068	(0.73, 0.99)

Posterior estimates of these residuals are plotted in Figures 6.14, 6.15 and 6.16 to illustrate the relationships between residuals. Figure 6.14 shows that students with high mean performance in 'standard' (100% examination) modules tend to experience greater losses or smaller gains from a 100% coursework regime rather than 100% examination than other students. Figure 6.15 shows a similar pattern holds for modules using mixed coursework and examination assessment. Figure 6.16 illustrates the strong positive correlation between the residuals associated with 100% coursework and mixed assessment.

Figure 6.14 Estimated student level residuals $\{\hat{v}_{consk}\}$ vs $\{\hat{v}_{allcswk}\}$

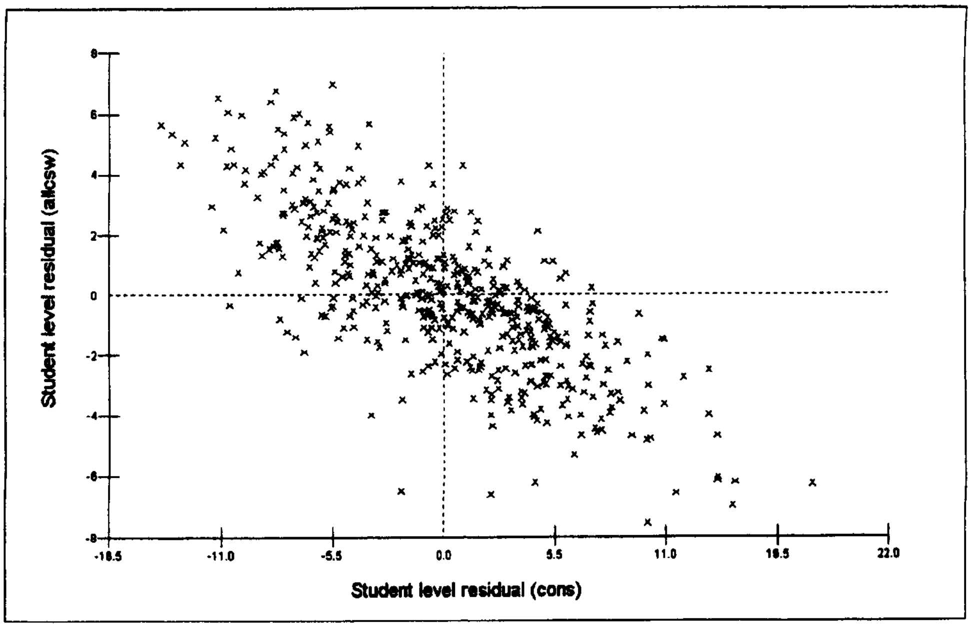


Figure 6.15 Estimated student level residuals $\{\hat{v}_{consk}\}$ vs $\{\hat{v}_{somecswk}\}$

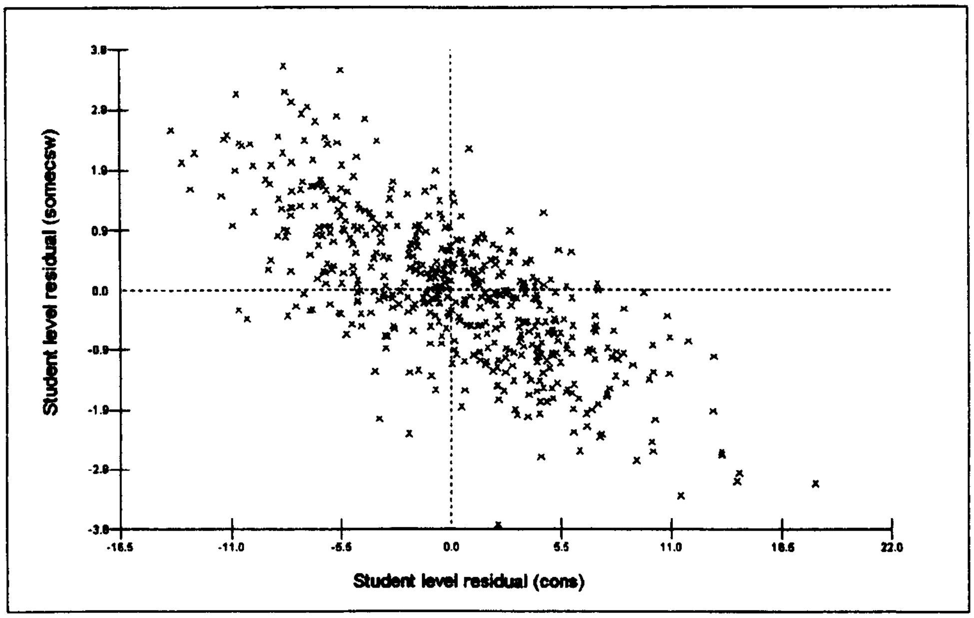
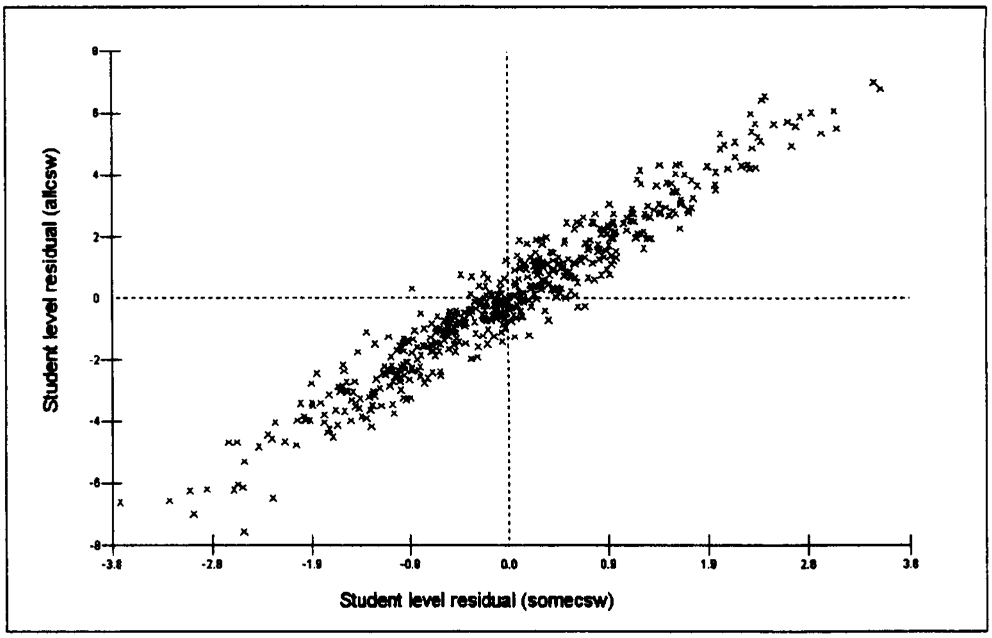


Figure 6.16 Estimated student level residuals $\{\hat{v}_{allcswk}\}$ vs $\{\hat{v}_{somecswk}\}$



The associations shown in Figures 6.14-6.16 do not imply that some students excel (relative to other students) in examinations while others excel in coursework assignments. In 100% coursework modules an individual's mean response is distinguished from the overall mean level of performance in similar modules by the sum of two student-level residuals $(v_{consk} + v_{allcswk})$ and similarly, in modules using mixed assessment by $(v_{consk} + v_{somecswk})$. The correlation coefficients for these quantities are:

$$\rho(v_{cons}, v_{cons} + v_{allcsw}) = \frac{\sigma^2_{vcons} + \sigma_{vcons/allcsw}}{\sqrt{\sigma^2_{vcons}\left(\sigma^2_{vcons} + 2\sigma_{vcons/allcsw} + \sigma^2_{vallcsw}\right)}}$$

$$\rho(v_{cons}, v_{cons} + v_{somecsw}) = \frac{\sigma^2_{vcons} + \sigma_{vcons/somecsw}}{\sqrt{\sigma^2_{vcons}\left(\sigma^2_{vcons} + 2\sigma_{vcons/somecsw} + \sigma^2_{vsomecsw}\right)}}$$

$$\rho(v_{cons} + v_{somecsw}, v_{cons} + v_{allcsw}) = \frac{\sigma^2_{vcons} + \sigma_{vcons/somecsw} + \sigma_{vcons/allcsw} + \sigma_{vsomecsw/allcsw}}{\sqrt{\left(\sigma^2_{vcons} + 2\sigma_{vcons/somecsw} + \sigma^2_{vsomecsw}\right)\left(\sigma^2_{vcons} + 2\sigma_{vcons/allcsw} + \sigma^2_{vallcsw}\right)}}$$

These represent the correlation between the mean marks achieved by the same students in modules using different types of assessment, other things being equal. The values of these correlation coefficients were calculated for each iteration in the chain and used to obtain estimates, standard errors and 95% confidence intervals. These are presented in Table 6.8 and the posterior estimates of the residuals are plotted in Figures 6.17 – 6.19.

Table 6.8 Correlation between student's individual means for three forms of assessment, standard students and modules

Correlation coefficient	Estimate	Standard error	95% confidence intervals
$\rho(v_{cons}, v_{cons} + v_{allcsw})$	0.80	0.040	(0.72, 0.88)
$\rho(v_{cons}, v_{cons} + v_{somecsw})$	0.97	0.011	(0.94, 0.99)
$\rho(v_{cons} + v_{somecsw}, v_{cons} + v_{allcsw})$	0.91	0.019	(0.87, 0.94)

There are strong correlations between an individual's mean performances in modules using all three kinds of assessment. The strongest correlation is between individual mean performances in modules using some coursework assessment and modules using 100% examination assessment, and the weakest correlation, shown in Figure 6.17, is between mean performance in modules in which assessment methods differ the most (using 100% coursework or 100% examination). Overall the conclusion is that, other things being equal, high achieving students perform well under any form of assessment, and their levels of achievement are the least affected by the choice of assessment method. It seems that methods of assessment have little impact on the relative order of students and more impact on the level of marks awarded, on the variation between students and on the consistency of the marks achieved within an individual's programme. In addition to these effects there is some evidence that lower achieving students are more affected by the choice of assessment method than higher achieving students, other things being equal.

Figure 6.17 Estimated student-level residuals $(\hat{v}_{cons})$ vs $(\hat{v}_{cons} + \hat{v}_{allcsw})$

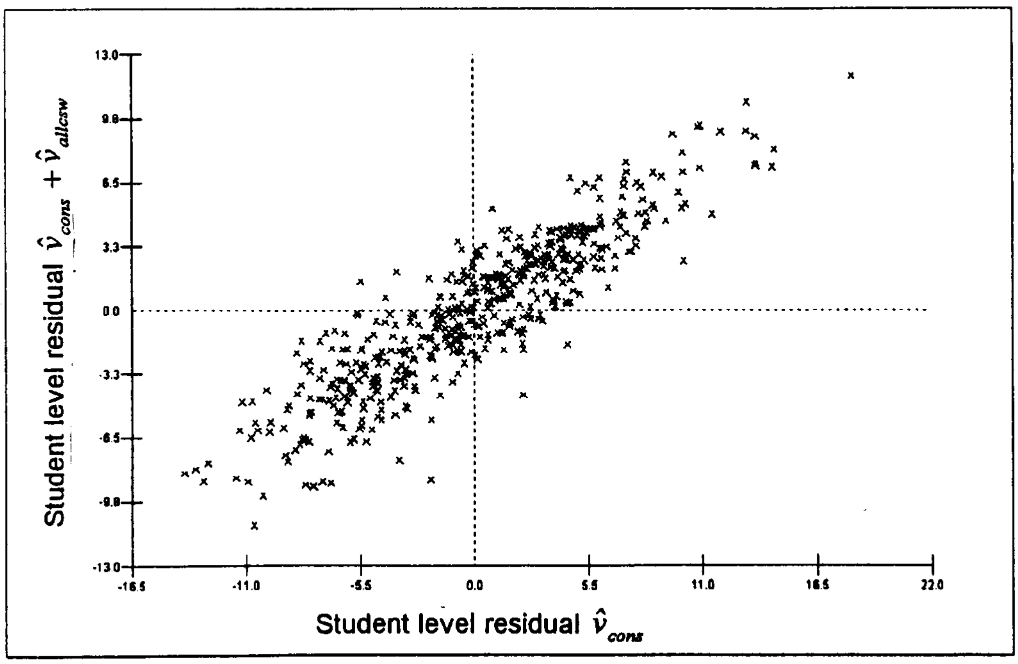


Figure 6.18 Estimated student-level residuals $(\hat{v}_{cons} + \hat{v}_{somecsw})$ vs $(\hat{v}_{cons} + \hat{v}_{allcsw})$

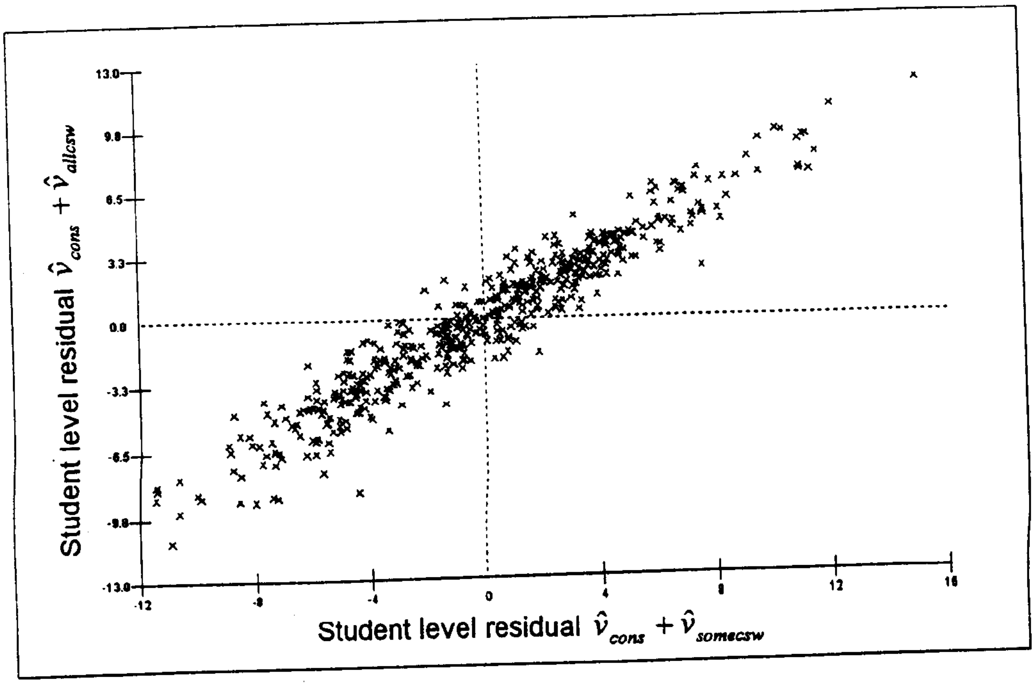
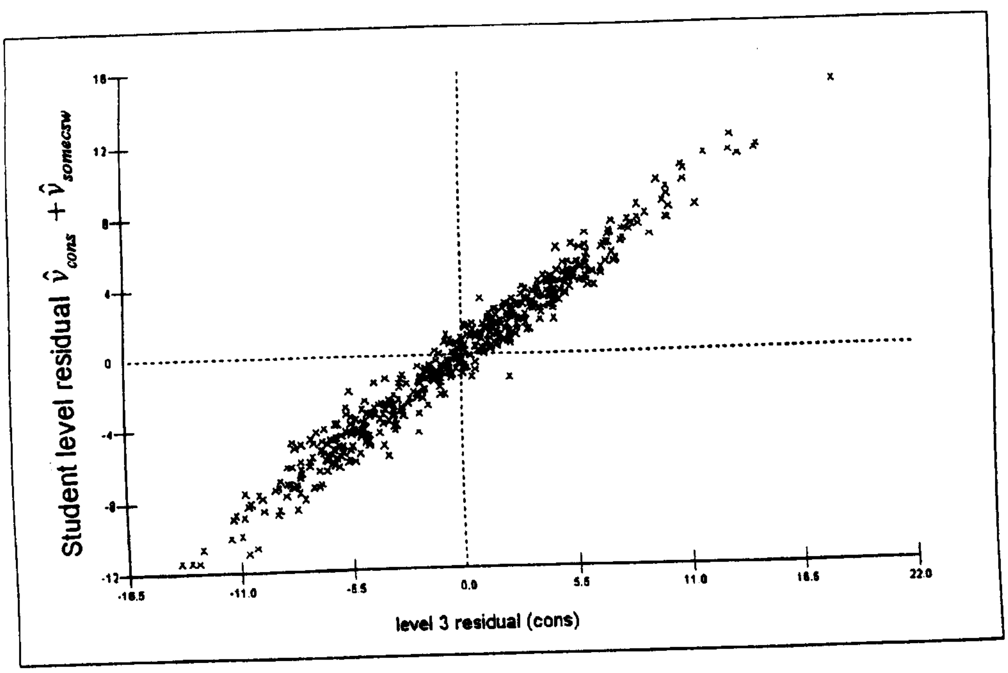


Figure 6.19 Estimated student-level residuals $(\hat{v}_{cons})$vs $(\hat{v}_{cons} + \hat{v}_{somecsw})$



6.8.1 Variations in Progress

In the final stage of the hierarchical analyses reported in chapter 4, the parameters β_{year1}, β_{year2} and β_{st2lin} were allowed to vary between students. No significant variation was found in β_{st2lin}, the parameter measuring term by term progress in years 2 and 3, but allowing the parameters concerned with year to year changes to vary at student level significantly improved the model and these random effects were incorporated in the cross-classified model reported in Table 6.2.

According to model (6.2), a 'standard' student, *k*, experiences a change in their mean marks between the last term of the first year and the first term of the second year of $-(\beta_{year1} + \beta_{term3} + 2.5\beta_{st2linear} + \nu_{year1k})$. Earlier it was reported that this change has an estimated mean of -0.579 (se = 0.457) marks, and estimated variance $\hat{\sigma}^2_{year1} = 5.92$ (se = 0.966). The variation between students (σ^2_{year1}) in this 'step' between the third and fourth terms is large enough for both substantial increases and decreases in marks to be relatively common: it is estimated that 26% of students (se = 6.3%, 95% confidence interval (14.6%, 39.1%)) will increase their mean marks by 1 mark or more in the transition from the end of the first year to the beginning of the second, while 28 % (se = 6.4% , 95% confidence interval (16.2%, 41.2%)) will experience a mean decrease of 2 marks or more during the same period.

Values of $\beta_{year3} + \nu_{year3k}$, individual *k*'s mean 'step' between second and third years, after allowing for the usual term by term improvement, have estimated mean $\hat{\beta}_{year3}$=2.01 (se =-0.531) and variance $\hat{\sigma}^2_{vyear3}$=7.32 (se = 0.972). An estimated 77% of students enjoy a positive 'step' between second and third years (se = 6%,95% confidence interval (64.0%, 87.6%). An improvement in mean achievement between the second and third years, in addition to progress made term by term, will not only have a positive effect on the average marks used to determine degree class but will also mean that the student is more likely to meet the criteria for classifying borderline candidates, as explained in section 3.7 .

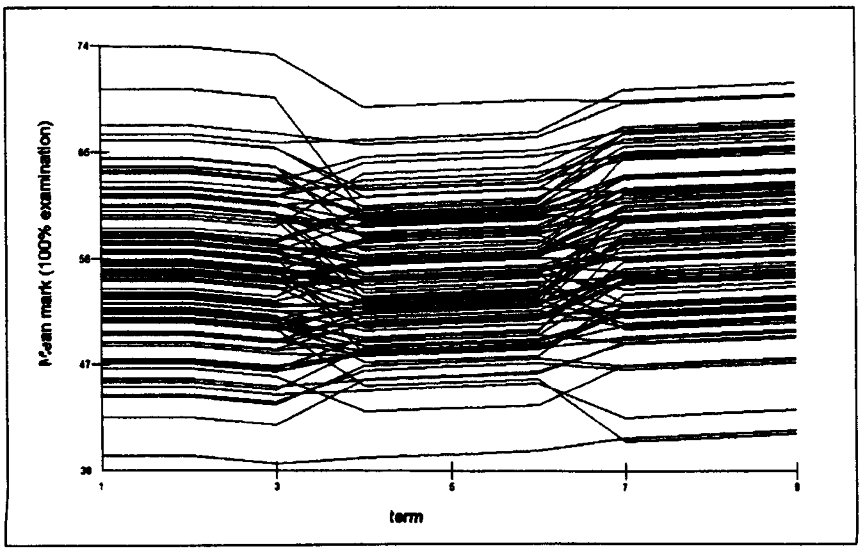
An individual's pattern of progress is weakly related to their response to different forms of assessment: as significant correlations were found between the student-level residuals associated with SOMECSW, ALLCSW and YEAR1. These are easier to interpret when considered in terms of the correlations between the change in marks between the end of the first year and the beginning of the second and the student's individual 'sensitivity' to different methods of assessment. The change in performance between the first and second year is positively correlated with the contrast between the students mean performance in 100% coursework and 100% examination modules ($-\hat{\rho}_{vallcsw/year1} = 0.39$,se = 0.153; 95% confidence interval 0.14, 0.62) and with the contrast between individual mean performances in modules using a mixture of assessment methods and in 100% examination modules($-\hat{\rho}_{vsomecsw/year1} = 0.6$, se = 0.153; 95% confidence limits 0.27, 0.85). The estimates of these correlation coefficients show that the students who gain the most from coursework assessment tend to improve between the end of the first year and the beginning of the second year, although the relationships are only moderate.

Having described variations in individual progress parameters, the next section is concerned with the overall patterns that this model of student progress leads to.

6.8.2 Individual Progress Charts

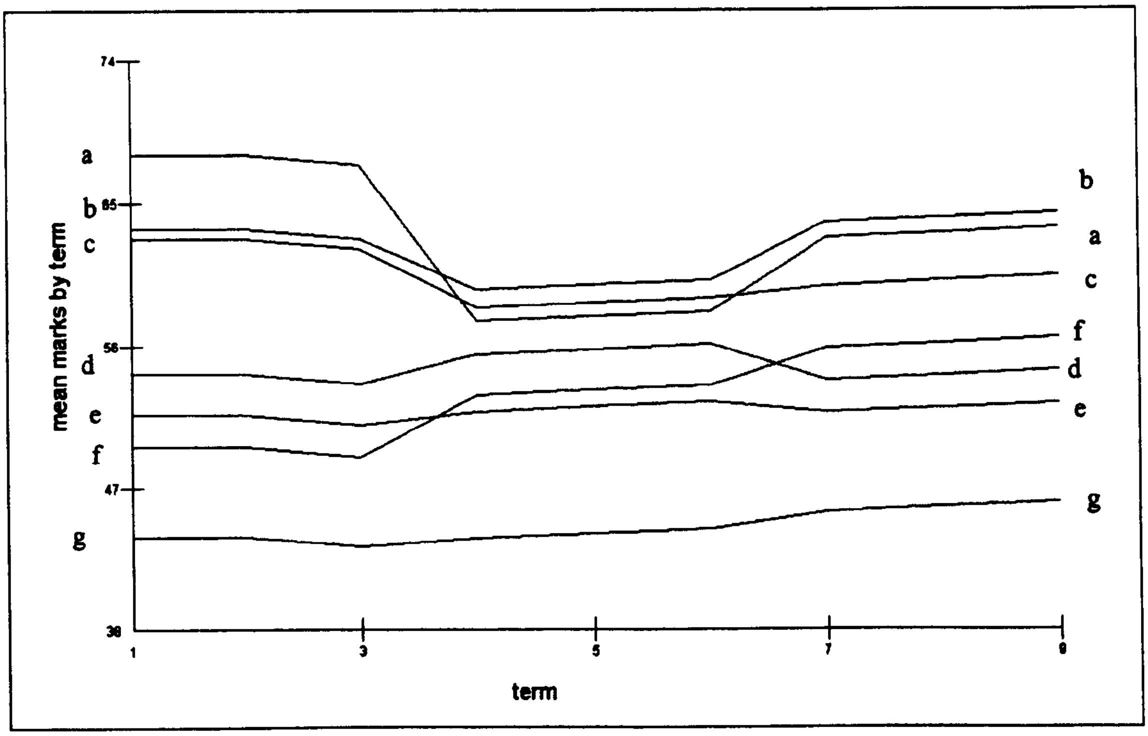
Figure 6.20 plots simulated 'progress charts' for 100 students belonging to the 'standard' or reference categories described in section 6.3, in 'standard' modules using 100% examination assessment.

Figure 6.20 Simulated mean marks, by term, for 100 'standard' students in 'standard' modules using 100% examination assessment



The most noticeable features of this graph are: the vertical spread of lines showing the variation between students' mean levels of achievement, the shifts in performance between academic years and the variations in the size and direction of these shifts shown by the crossing of the lines between adjacent years. Aside from changes between academic years, term by term changes are relatively small, and there is no evidence of important variation between students. Similar plots for modules using other forms of assessment show the same patterns but with less dispersion and higher intercepts. The large number of lines in Figure 6.20 makes it difficult to inspect the variety of progress charts produced. Seven were chosen to represent as wide a variety of shapes as possible: these are shown in Figure 6.21 and illustrated the variety of potential shifts in mean marks between academic years and their effects on students' levels of performance and relative positions.

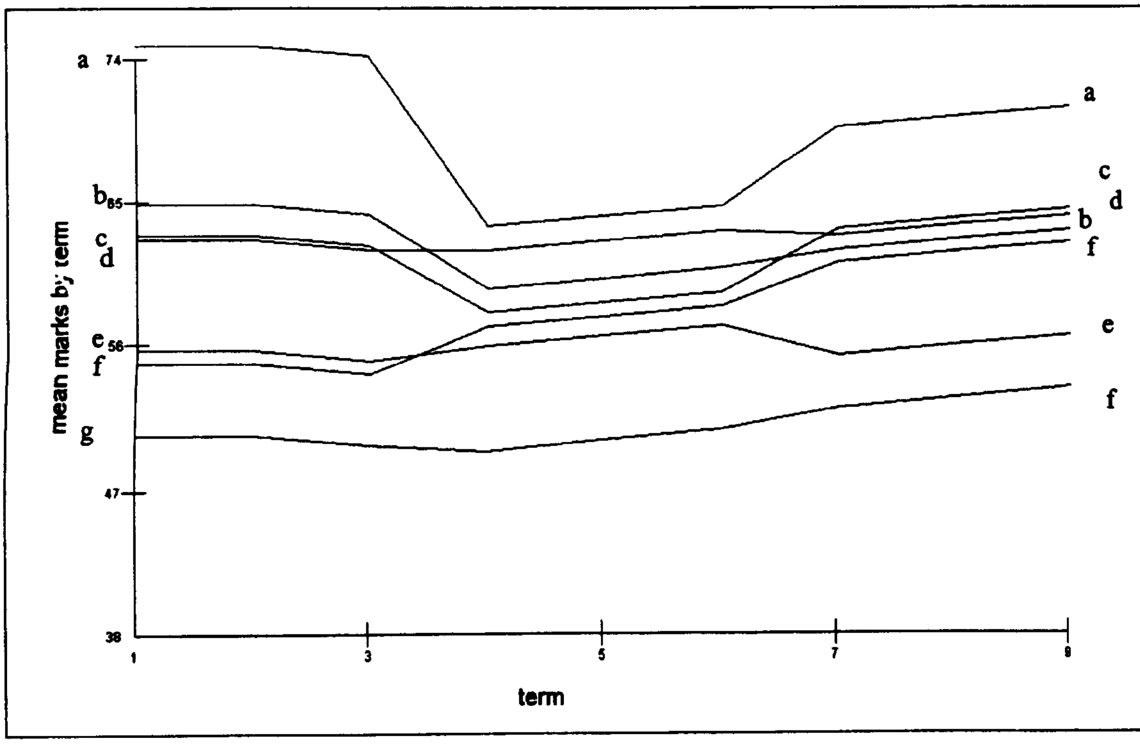
Figure 6.21 Selected progress charts, mixed assessment modules



While some lines (e and g) are relatively flat, others (such as a and f) display substantial shifts in performance and by the end of the three year period, there have been a number of changes in the rank ordering of the students' termly mean performance.

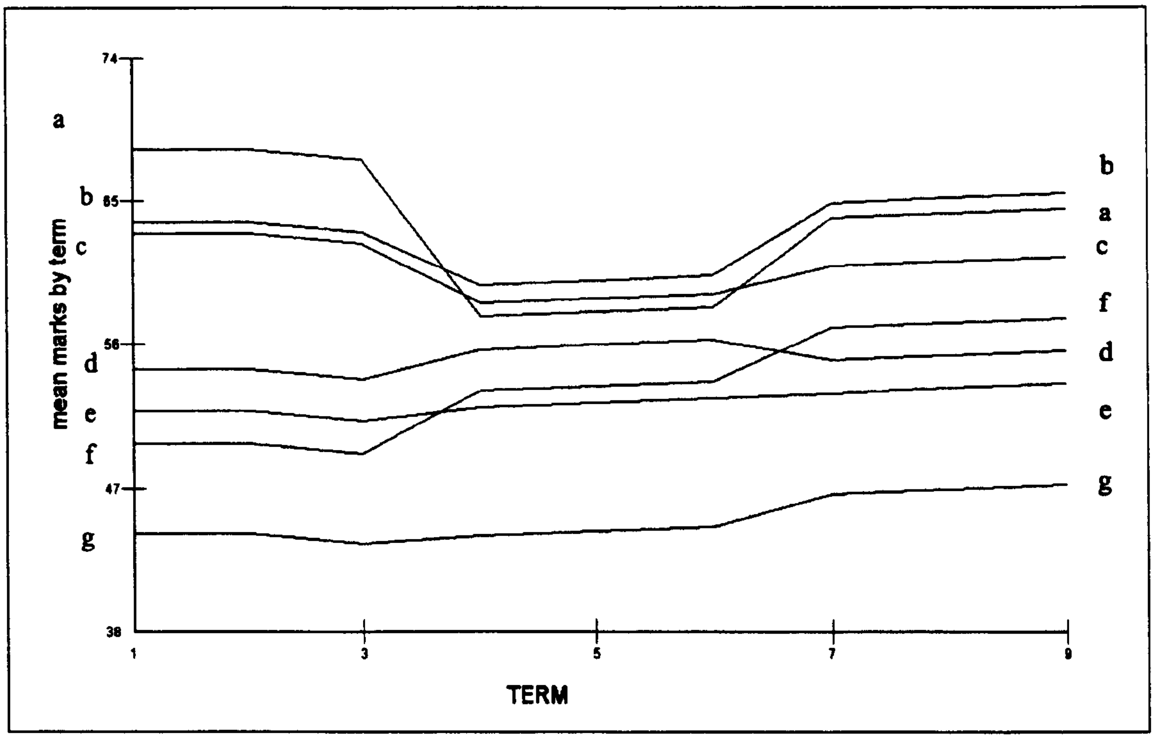
The diagram would be altered if the selected (simulated) students had belonged to other categories (mature or having a parent in a professional or managerial occupation), as these groups were identified as having significant differences from 'standard' students in some aspects of their performance. The effect on patterns of progress are shown in Figures 6.22 and 6.23.

Figure 6.22 Selected progress charts : mixed assessment, mature students



In Figure 6.22 , the characteristics of mature students have been added to the individual progress charts selected for Figure 6.21. The new shapes reflect the additional variation between mature students, here leading to different relative initial positions for the selected individuals, combined with higher mean marks and flatter progress in stage 2. The tendency for individuals to change position and for their mean marks to rise or fall between academic years is the same as for traditional students.

Figure 6.23 Selected progress charts: mixed assessment, professional/ managerial background



In Figure 6.23, the characteristics of students from a professional or managerial background were imposed on the individual progress charts selected for Figure 6.21. The new patterns of progress obtained are similar to those shown in Figure 6.21, but show greater progress between the second and third years.

6.7 Summary

This chapter has presented the results, including diagnostics output, for the MCMC estimation of a multi-level model based on a cross-classified structure. The computer time taken to fit this model was many times greater than the time required to execute earlier hierarchical models fitted by IGLS, as a result of fitting the cross-classification between students and modules. The output from the MCMC analysis confirmed the results based on variance components models, showing the cross-classified structure to be more appropriate than the hierarchical structure which ignores the clustering of students' entries within modules.

Estimates of parameters describing the effects of background, progress and module characteristics on the mean level of achievement, on the consistency of students' performances and on the variation between students. All of these effects have an impact on students' records and on their degree classifications, either directly, by affecting the mean marks achieved in modules used to classify students' degrees, or indirectly, by influencing whether a student meets the criteria for the upgrading of students whose degree classification is borderline. Some of these results confirm published findings discussed earlier, in chapter 2, while others provide completely new information. The use of a multilevel, cross-classified model, including complex variation at level 1 and the analysis of full records represents a new approach to studying undergraduate achievement. The next chapter will evaluate the findings in the context of published research studies and review what has been learnt from the application of the techniques described in chapter 5 .

Chapter 7

Conclusions

7.1 Introduction

This chapter reviews the research findings presented in earlier chapters.

One feature of the research presented here is that it is based on analyses of the academic records of a particular sample of students from a single modular scheme. Section 7.2 discusses how the selection of this sample may have influenced the research findings and whether conclusions based on analyses of these students' academic records can be extrapolated to other contexts.

Chapter 1 listed the aims of this thesis. Briefly, these were

- to provide a model for the analysis of undergraduate achievement within modular degree programmes
- to re-evaluate the findings of earlier research in higher education while controlling for a wider selection of explanatory variables
- to model patterns of student progress
- to study factors affecting the variation in performance between students
- to study factors affecting the consistency of a student's performance.

Sections 7.3 – 7.7 of this chapter will review the findings of earlier chapters and show how the aims have been achieved.

Finally, the production of new information about undergraduate achievement raises new questions about undergraduate achievement that could be investigated in other studies. The application of relatively new statistical techniques to data from a modular degree course leads to questions of how these techniques might be used to investigate other questions relating to undergraduate achievement or be applied to other contexts. These issues are briefly discussed in section 7.8.

7.2 Sample Selection/Extrapolating the Results

The research presented here is based on analyses of data for a sample of students selected from those studying within a modular degree programme at a single institution. This section is concerned with how the selection of these students may have influenced the results of this research and whether the findings of research carried out within one modular scheme can usefully be applied to others.

The selection of students who graduated from the Modular Degree Programme at Oxford Brookes University within three academic years excludes several categories of students. These include full-time students who failed to complete their programme or withdrew from the course, students studying part-time, students transferring in or out of the Modular Degree Programme from other universities and students who enrolled on full-time courses taking four years or including an extended period of assessed practice. Concentrating on students with records covering three academic years simplified the process of identifying a suitable form for fitting individual progress charts. The information about progress, though limited to describing the records of students graduating on time, is original and provides a baseline against which the progress of other groups can be studied in future. Even though it seems reasonable to restrict attention to the performance of students following a course of the same duration, the effects of this selection need to be considered.

The study of a sample of graduates means that the selected students are the more successful of those who enrolled for a full-time, 3-year degree course. Amongst these students, variation between students will tend to be lower and estimated mean levels of achievement will tend to be higher than the corresponding figures for all full-time students who enrolled on three-year degree courses at the same time. The 'progress charts' and parameter estimates obtained are therefore likely to provide an optimistic view of student progress.

On the question of whether research carried out at one institution can be applied to others, it seems reasonable to suppose that the extent to which the conclusions of this research can be extrapolated to students on other modular first degree courses will depend on the similarities between these courses and the Modular Degree Programme at Oxford Brookes University. The research literature demonstrates the consistency of some effects across similar institutions: for example, Yorke et al (1996) demonstrated consistencies in subject differentials across a number of modular schemes in 'new' universities. There are also instances where findings based on one institution have been used to explain patterns of achievement in others or across the whole of the higher education sector. For example, Gibbs and Lucas's (1997) paper, analysing data for the Modular Degree programme at Oxford Brookes University and concluding that '*it is easier for students to gain good marks in modules with higher proportions of coursework assessment and easier to get good degree results in subject areas which use higher proportions of coursework assessment*', has been used to explain changes in the proportions of first class and upper second class degrees awarded across the whole of the higher education sector (Elton, 1998; HEQC, 1997).

While modular schemes operate within similar frameworks, details such as the mark or grade system, the sizes and levels of modules, progression rules and classification systems vary. It seems likely that while some of the effects identified in earlier chapters may be characteristic of modular higher education per se, others may depend on particular features of a modular scheme, such as the scale on which achievement is measured or the degree classification system, and the way that students respond to these features. Hence while there may be similar relationships between assessment method and performance in different schemes, the values of the relevant parameters are likely to differ.

Differences between modular schemes in some features may lead to patterns of achievement which differ from those identified here. Chapter 2 provided an example showing how the implications of interim results depend on the system of aggregation and classification in use. This example suggests that different schemes may evoke different responses to feedback on performance from students whose records are similar. If so, then patterns of progress may differ between institutions as a result.

Whether or not there are sufficient grounds for using the findings of this research to draw conclusions about modular programmes in other institutions, the findings are relevant to other modular degree programmes in identifying questions that can be addressed by research. For example, the finding that female and/or mature students perform more consistently than others at Oxford Brookes University raises the question of whether the same patterns occur in other modular programmes and if not, then how this could be explained by institutional characteristics. One such question is whether the consistency of students' performance, or differentials in consistency, are influenced by the extent to which the aggregation system rewards consistent performance.

New research questions generated by the current findings will be discussed in more detail in section 7.8. The next sections are concerned with showing how the aims listed in chapter 1 have been met.

7.3 Methodological approach for studying undergraduate achievement in a modular degree course

The analyses described here differ from those in earlier studies of undergraduate achievement, in that students' academic achievement is measured in each module entry within their degree programmes. The marks awarded in each entry are not independent: each student contributes a number of responses on each of a number of occasions and some of the marks awarded to different students were awarded within the same module. It is important to use a model that represents this structure, taking into account the lack of independence between module entries. Such models provide efficient estimates of regression coefficients, appropriate standard errors and the opportunity to model random effects and complex variance functions (Goldstein, 1995). This was illustrated in the results obtained for the final model, where parameters associated with some module characteristics became non-significant once the cross-classified structure was added.

The value of being able to model variances is demonstrated by the findings based on estimated variance parameters. Some of these findings describe variation between students, providing information of value to lecturers, assessors and providers of learning support within the Modular Degree Programme at Oxford

Brookes, who need to consider the needs of the full range of students. The potential danger of creating stereotypes based on averages is illustrated by the finding that although mature students have mean marks significantly higher than those of traditional students, in terms of achievement as well as experience, they are a more diverse group than those entering the course at a more traditional age.

Modelling variation at level 1 recognises that student records do not consist of uniform results and that this inconsistency may be related to student or module characteristics. An earlier section (4.2) established the difficulty of constructing a rationale for standardising raw marks. Level 1 complex variance parameters related to subjects, with relevant fixed parameters, were essential in controlling for disciplinary differences in the distribution of marks awarded within modules. Estimates of these parameters show how the marks awarded within modules, controlled by different subject examination committees, lead to subject differences in the distribution of graduates' degree classifications even though subject examination committees have little control over the largely automatic classification system.

Initially, a series of hierarchical models was fitted to the data to determine which factors influence the marks awarded within a student's programme and how these effects should be represented in the model. In these analyses, maximum likelihood estimates of the model parameters were obtained using IGLS. Chapter 5 discussed alternative structures and presented analyses of test data sets. These analyses showed that fitting the more complex structure was worthwhile, but the number of units involved in the cross-classification was too large for the cross-classified model to be fitted using the same estimation technique (IGLS) as was used to fit the hierarchical models and so alternative estimation methods were explored. Chapter 5 described MCMC estimation techniques capable of fitting a cross-classified model that also included complex level 1 variances and random effects: these techniques were used to fit the final model as described in chapter 6.

The use of IGLS to obtain maximum likelihood estimates at one stage and MCMC estimation at another, had a practical advantage which could be of use in other studies: as the hierarchical model was built up, effects were selected and rejected at each stage by comparing models with and without each parameter. Even though this required a large number of models to be fitted, the overall process of selection was relatively quick as, using IGLS, estimates were computed within seconds. Once the choice of independent variables for each part of the hierarchical model was established, only a single cross-classified model remained to be fitted using MCMC estimation. MCMC estimation is much slower, but is capable of fitting a cross-classified model to a larger dataset.

The advantages associated with the approach used here are: appropriate standard errors are obtained for the parameter estimates, students' achievements are analysed in detail, individual models of changes over time can be fitted and variation is modelled at different levels. These features, combined with the use of a model that reflects the true structure of the data, mean that the research provides a model for future studies of undergraduate achievement within modular degree courses.

7.4 Evaluation of Current Findings in the Context of Previous Studies

The literature review in Chapter 2 highlighted the limitations of published research studying undergraduate performance: although studies in higher education have examined several performance-related factors, our ability to evaluate the effects of these factors is limited by the small number of multi-factor analyses available and the use of degree classification as a measure of achievement. There were also gaps in the research: no published studies were found *describing students' progress* during the course of their degree, the use of fixed effects models means that very little evidence is available on variation within sub-groups of students and, in spite of the popularity of modular courses there appears to be no published evidence relating to the consistency of performance of individual students within such courses. In the analyses reported here, achievement is measured in terms of marks rather than classes and a large number of student and module characteristics, and interactions between them, are included as explanatory variables, so that the limitations of the methods used in previous studies are avoided. This section compares selected findings of this research with relevant results from earlier studies of achievement in higher education. The modelling of student progress and use of random effects models lead to new findings; these will be reviewed in sections 7.5 and 7.6 respectively.

7.4.1 Sex differences

The studies reviewed in chapter 2 had explored the effects on academic achievement of undergraduates' age, sex, entry qualifications, class size and assessment methods. Section 2.3.1 found that some studies had reported that women were less likely than men to be awarded degrees in either the highest or lowest degree classes. Having only limited control over the effects of other factors disadvantaged several studies investigating this topic. Tomlinson and MacFarlane (1995) showed that at least some of the differences between the distributions of male and female graduates by degree class could be explained by differences in their choices of degree subject. The analyses reported in chapters 4 and 6, avoiding the methodological limitations of the earlier studies, found the mean marks achieved by male undergraduates to be lower than those achieved by female undergraduates. While developing the final model, no evidence was found that the between student variation was greater amongst males, however the variation in performance *within* male graduates' records was found to be higher that that within female graduates' records. Under a classification system similar to that used by Oxford Brookes, this greater variation could produce a tendency for male graduates to have more diverse degree classifications than female graduates, other things being equal. Amongst students with a low mean level of achievement, male students' greater variability in performance would be associated with a higher risk of failure. With fewer module credits, male students would be less able to accumulate sufficient module credits to graduate with honours or to benefit by excluding their lowest marks from the calculation of degree class. For male students with higher mean levels of achievement, the higher level 1 variation is a benefit: these students gain more from the selection of their best results to determine degree class and are more likely to meet the criteria for upgrading borderline cases.

It is difficult to draw conclusions when comparing the findings on sex differences in academic achievement between this and other research studies, as this requires comparisons to be made between different generations of graduates. The cohort studied here graduated in 1997, but many of the samples used in earlier research consisted of students who had graduated in the 1980 or earlier: even Johnes' analysis, published in 1992, was based on a cohort graduating in 1980. Comparing the approach used here to the methods used in earlier studies of sex differences in higher education, we see that information at module level shows how differences in degree classifications are produced, as the result of sex differences in means and variances, within a modular system.

7.4.2 Age

The more recent studies of the academic achievement of mature students, largely carried out in 'new universities' and reviewed in Chapter 2, found that mature students tend to be more successful than those in the traditional age group. In the analysis described in chapter 6, the mean marks achieved by mature students were an estimated 3.49 (se=0.621) marks higher than for similar traditional students, a very large difference, given the grade bands used within the system. As with gender differences, differences between age groups were more complicated than a simple mean difference. Interactions were found between age and assessment method and between age and the linear element of student progress and age was also a factor contributing to complex variation at both student and module entry level. Although mature students achieved higher mean marks than traditional students and performed more consistently, it would be wrong to imagine them to be a uniformly successful group, as the complex variance parameter showed that between student variation amongst mature students was significantly higher than amongst traditional students. As with other factors, the use of a multi-level model with random effects and complex variance functions leads to a detailed picture of how students in different age groups differ.

7.4.3 Entry Qualifications

Problems with missing data and inconsistencies arising in the entry qualifications recorded in the CSMS were discussed in section 3.9.2 which explained that the minimum entry grades for the students' chosen degree were used as a measure of students' prior educational achievement. This means that although some allowance was made for educational achievement prior to entry, it is not reasonable to use the current analyses to measure the impact of prior attainment on undergraduate performance. The ability of similar studies to provide information on this topic may improve as systems for recording information have been updated in order to meet the requirements of HESA.

7.4.4 Method Of Assessment

The main research evidence on the effects of assessment methods on achievement in higher education is provided by Gibbs and Lucas's (1997) study of module averages at Oxford Brookes University. This showed that module averages tended to be higher in modules using coursework assessment. Other research concerned with assessment seems to be mainly concerned with studying attitudes to different forms of assessment, rather than measuring their impact on achievement. There seems to be no published research concerned with the potential for assessment methods to influence differences between sub-groups of students in achievement.

Longitudinal data recording students' experience of assessment throughout their programmes showed that the percentage of marks available in examinations varied within students' programmes and from one student's programme to another. Given these variations, the most appropriate way to study the effects of assessment methods is by analysing the marks achieved by students in each module and this has allowed the effects of assessment methods on undergraduate achievement to be studied in greater detail than in previous studies. The final analysis presented in chapter 6 found that assessment methods had significant and substantial effects on students' marks in the form of main effects and interactions, fixed and random effects and complex variance terms. These results confirm that the tendency for coursework to raise average marks, described by Gibbs and Lucas (1997), remains after controlling for other factors. Other findings related to assessment methods provide new information, describing the effects of assessment methods on: the differences in achievement between subjects and between mature and traditional students, between-student variation and the consistency of individual students' performance. The model represents the effects of assessment methods as varying between students, a more realistic approach than one that assumes that the impact of different assessment methods is the same for all students. The relevant parameter estimates showed substantial variation between students in the effects of different forms of assessment, something which needs to be taken into account when changes in assessment practices are considered.

Within a modular programme, effects of assessment methods on student's achievement raise concerns about 'standards' and questions of 'fairness to students'. Discussions of 'standards' are concerned with whether coursework assignments are 'easier' than examinations or are marked more generously. Questions of fairness arise because students' degree awards should reflect individual achievement rather than the assessment methods experienced in their programme. The implication is that if the design or marking of assessments were adjusted, then these difficulties regarding 'fairness' and 'standards' would be resolved. The conclusion, in chapter 6, that assessments have different effects of different students is important to this debate, as it shows that no simple adjustment can be made to equate two forms of assessment as while one form of assessment may produce higher marks on average than another, there will be students for whom the reverse is true.

7.4.5 Other Factors

Other findings include those relating to class size and subjects. These two factors are particularly important as modular schemes are associated with interdisciplinary programmes and economies of scale (HEQC, 1996).

Most of the existing evidence of the effects of class size on student achievement in higher education is based on a series of studies carried out at Oxford Brookes University (Fearnley, 1995;Lindsay and Paton-Salzburg, 1987; Gibbs and Lucas, 1997). These analyses share common limitations, being based on data aggregated at module level and failing to control for variations between modules in the characteristics of the student intake. This is problematic, as Chapter 3 showed that each student experienced a variety of class sizes during the course of their degree, with class sizes tending to fall during the course of students' programmes, while at the same time, students become more experienced. Analyses that fail to take account of students' experience therefore risk attributing improvements due to progress to the reduction in class size. This problem was avoided by using multilevel models that allowed the effects of class sizes on students to be measured directly and after controlling for other performance related factors. Once other factors had been taken into account, class size did not appear to influence student's performance. Potential explanations for this finding are that assessors are able to match their assessment criteria to what can be achieved in modules with a given enrolment, or alternatively that strategies for teaching classes of different sizes are successful.

In the final model described in chapter 6, several parameters represented the impact of subjects on students marks, including fixed effects and complex variance at level 1, so that estimates of the effects of other factors were adjusted for the mark distributions characterising different subjects. This approach represents a step forward in terms of controlling for the effects of subject differences in studies of achievement, and in the study of subject differences *per se*. Chapter 2 reported that previous studies of the mark distributions in different subjects failed to control for the clustering of marks within modules and/or individuals. It was also noted that studies of achievement in higher education have controlled for the effects of differences between subjects in mark distributions in a limited way. The analyses found significant and substantial differences between subjects in the mean marks achieved and in level 1 variation, illustrating the difficult questions of comparability between subjects that are raised by modular schemes.

Some interactions between subjects and assessment methods were found: further research might determine whether these were explained by different weightings given to coursework and examination or by differences in the nature of the assessments set.

7.5 New Information Concerning Student Progress

As there are no published longitudinal studies of undergraduate achievement, the findings related to student progress are all new. The analyses in chapter 6 showed that some terms of a student's degree programme are associated with lower mean marks and/or higher level 1 variation. Either of these effects increases a student's risk of failing a module. An advantage of modular schemes is that students have the opportunity to change their programmes or modes of study to regain a viable position, but in some cases, these strategies may fail or be rejected. The typical pattern of progress identified in chapter 6, identifies the fourth term as the point within students' programmes at which mean marks are lowest and the first year as a period in which students perform less consistently than in other years. At these times, students are, in academic terms, at their most vulnerable. This information could be used to assist the development of a system of support designed to avoid disruptions and withdrawals.

In the first year, mean marks did not appear to vary between terms, but a shift in mean marks, varying between students, was identified between the first and second years. This shift coincides with the students' progression from basic to advanced modules and the possibility that their results will contribute to their degree classification. Relevant parameter estimates and the 'progress charts' plotted in section 6.8.2 showed that a wide range of changes, both positive and negative, can occur at this point in a student's degree. This raises questions about what determines a student's change in mean marks on entering the second year of their course. There was some evidence that this change was partly explained by the student's responses to different forms of assessment: this and other potential explanations could be the subject of further research. The results suggest that some students need support in the transition between the first and second years of their degree. If further research could identify factors influencing the size of the 'step' in performance between the first and second years, support could be effectively targeted on students who are most at risk of failing modules.

For a typical student, progress during stage 2 consisted of two elements: a small term by term increment, apparently confined to students entering the degree course at the traditional age, combined with a 'step' in performance between the second and third years. A characteristic of the Oxford Brookes Modular Degree Programme is that modules are offered at only two levels, so that modules taken in the third year are at the same level as modules taken in the second year. If there were no flexibility about whether to include modules in the second or third year, then changes in mean marks during stage 2 could be produced by changes in the standards of assessment applied in modules taken in different terms or years within a degree programme. As a degree of flexibility exists, allowing students to exercise some choice as to when, in the second or third years of their programme, to take certain modules, it is reasonable to assume that at least some of this progress can be attributed to the students becoming more experienced, more motivated, developing skills and building on their knowledge as they acquire more module credits and obtain regular feedback on their performance. The average term by term improvement is quite small (95% confidence interval (0.01, 0.59) marks), but the importance of this element of 'progress' is reinforced by the contrast between term by term progress for students in different age groups. Traditional students achieve greater term by term improvements than mature students during stage 2, the 95% confidence interval for this difference in term by term gain being (0.14, 0.57) marks. This finding would seem to suggest that improvements in transferable skills, motivation, and effort, rather than subject expertise, explain the linear element of traditional students' progress. Other than the interaction by age, there was no evidence that the term by term increment varied between students.

The third element of student progress is the 'step' between second and third years. At this point there is no change in the level of modules taken and the shift is measured after allowing for the usual term by term improvement. As with the progression from the first to the second years, the size of the change and its direction, varied between students. One of the most noticeable features of the graphs of simulated progress charts was the crossing of lines showing that students' mean marks can move up or down, from one academic year to the next, in steps of differing sizes. Students from a professional or managerial background achieved larger improvements, on average, between second and third years than other students. The simulated 'progress charts' shown in section 6.8 showed that some students experience substantial changes in mean marks as they enter the final year of their degree. This phenomenon has not been mentioned in the research literature, although degree classification systems such as the one used at Oxford Brookes, are designed to reward students whose performance improves during their final year.

7.6 New Information Arising from the Random Part of the Model

Further original findings are those relating to variation. The final model represents the variation in marks as occurring at three levels: between students, between modules and at the level of individual module entries. Variation between modules was represented by a single parameter, σ_u^2; it was important to include this in the model, in order to recognise the clustering of module entries within modules. The variance matrix Ω_v was used to allow the effects of some factors to vary between students and to test for complex variance effects. Early analyses established that the between student variance was the same for male and female students, an important finding given some previous studies have found that degree classifications are more dispersed for male and female graduates. In the final model, estimates of Ω_v indicate greater variation between mature students than between students in the traditional age group. Students' ages also contributed to variation at level 1, where the level 1 variance was lower for mature students. The fixed part of the model showed that mature students, other things being equal, have higher mean achievement than younger students. Important additions to this information are provided by the random part of the model, which shows that although, as individuals, mature students perform more consistently than younger students, as a group their performances are more diverse.

The typical progress charts described in section 6.7 provide new evidence describing student progress within a modular degree programme. Student level variance parameters added to this evidence by showing how the 'steps' in achievement varied from one student to another.

Further examples of original findings are provided by the findings concerned with the effects of assessment method and discussed earlier in this chapter. These show that changes in mean achievement are just one aspect of the impact of assessment methods on students' marks. The level 3 variance parameters showed that the effects of assessment methods were not constant, but varied between students, to the extent that, for an estimated 20% of students (95% confidence interval (10%-32%), coursework assessment is associated with *lower* mean marks than examination assessment. This finding is particularly important since it means that changes in assessment practices will not affect all students equally.

No published studies have investigated factors influencing the variation in marks achieved within a student's programme, and yet this variation has important implications. A discussion, in chapter 2, of degree classification systems highlighted the effect of variations in student performance on degree class. Prospective employers and other interested parties may make judgements about a candidate's reliability based on a transcript showing their achievements in detail. Before graduation, variations in performance can lead to failed modules, with potential disruptions to their programmes and workload. All the findings related to level 1 variations therefore have important implications.

This section has highlighted some of the new information obtained from the random part of the model; in some cases these findings raise new questions and some of these are discussed in *the next section.*

7. 7 Future Research

The research described here can be extended in several ways.

First, the methodological approach used here can be applied to data for other students who have studied within the Modular Degree Programme at Oxford Brookes. Increasing the number of students in the sample would raise the numbers of module entries per module, so that the variation between modules could be investigated in a similar way to the modelling of variation between students in the analyses presented here.

The selection of the sample could be extended to include students who: transfer in or out of the course to other institutions, enrol as part-time students, enrol on sandwich courses or four year degree courses, study for degrees that include a substantial period of assessed practice, whose programme includes a period of temporary withdrawal, who extend the period of their degree as a consequence of failed modules or who leave the course without the award of a degree. Widening the selection of students would allow the effects of the factors studied to be measured for the whole population of students rather than those who achieve a traditional 3-year award on schedule. Modular degree courses facilitate changes between full and part time study, between institutions and between periods in education and other life courses. By studying the records of a wider selection of students, variations in the patterns of progress of students following different modes of study and over different time scales could be studied.

As student withdrawal is a concern in higher education; further analyses could investigate *patterns of progress* leading to withdrawal, with the objective of being able, in future, to identify those individuals most likely to leave without completing their programme.

Secondly, the approach used here should be applied to the records of students within other modular programmes to see whether the effects identified here appear in other modular degree courses. A larger analysis might draw on data from a number of courses or institutions. This would mean dealing with questions of comparability and the use of different measuring scales, but if these could be resolved, many interesting questions could be investigated. The courses or institutions would introduce an extra level of analysis, with the student/module structure nested within the new units defined by institutions. This would enable the impact of characteristics of institutions or modular programmes on progress charts or assessment differentials to be explored.

Thirdly, the current findings could be explored in greater detail in studies seeking to explain some of the findings reported here. For example, future studies might seek to explain why individual students respond differently to different assessment methods and the transitions between academic years or why students from a professional or managerial background typically made greater progress in their final year than other students.

Finally, the methodological approach used here should be applied to other contexts. The advantage of using statistical models that accurately reflect the structure of the data have been known for some time. The research presented here has shown that newly developed estimation techniques allow multilevel, cross-classified models to be applied more widely than has previously been possible.

Studies taking advantage of these new developments would potentially lead to valuable findings such as those described earlier. For example, in higher education, these techniques could be used in a system-wide study to examine factors affecting students' completion and/or degree class in first degree courses, with students cross-classified by university and previous educational establishment. This cross-classification would involve large numbers of units, but could be fitted using the approach used here. The facility to model the effects of student and institutional characteristics and to define the variation between students, universities and previous institutions attended in terms of explanatory variables would allow a useful and detailed study to be made of factors affecting students' ability to succeed in the transition from secondary to higher education. More generally, the methodological approach described in chapter 5 could be applied to other contexts in which longitudinal data is collected from individuals according to one hierarchical structure, while the observation or production of the those responses corresponds to the 'delivery' of a service within another hierarchy. In this research, the first hierarchy corresponds to the hierarchy of assessments within students' programmes, such that the level 1 units within the hierarchy correspond to the level 1 units within the hierarchy of assessments within modules 'delivered' by the modular course. In health care, an example of such a structure might be the longitudinal assessment of patients, cross-classified by the team/unit/carer responsible for their care or treatment or perhaps by the technician assessing them on each occasion. Similarly, longitudinal studies of salaries or satisfaction with work might cross-classify individuals by households and by employers. In both cases, large numbers of units would be involved in the cross-classification and the opportunity to model variation or to fit individual 'growth' curves would be useful and could be achieved using the MCMC estimation techniques described earlier, with models based on simpler structures being used to select which variables and effects to include.

7.8 Conclusions

This chapter has shown how the research presented here has advanced the study of undergraduate achievement. Using a more sensitive measure of achievement, and controlling for more explanatory variables than earlier studies, this study was able to re-assess the effects of some performance-related factors. The effects of some of these factors, such as coursework assessment, were confirmed, but the effects of other factors, such as class size, disappeared after controlling for other variables.

Analysing marks for modules within students' programmes made it possible to study achievement longitudinally, using a multilevel model that incorporated random effects and a complex variance structure, leading to original findings presented in chapter 6. Several of these findings raise new questions that could be explored in future studies.

The marks in students' records measured the individuals' achievement on several occasions and in each module in their programme, and many of the programmes had modules in common. This nesting of module entries within student programmes, coupled with the grouping of students' module entries within modules is characteristic of assessment within modular programmes. A number of practical difficulties were encountered in fitting a model based on this structure. These difficulties were resolved by the use of MCMC estimation using recently developed techniques enabling random effects and complex variance structures to be fitted within a cross-classified multilevel model. This approach provides a model for suture studies of undergraduate achievement within modular systems and can be used to extend the current research to cover both a wider range of students within the Modular Degree Programme at Oxford Brookes University and students in other modular programmes.

REFERENCES

Adams, R. (1996) Changing external examiner practice to meet the need of a modular framework, in : HEQC(1996) *Modular Higher Education in the United Kingdom: in focus*, London, 108-110, HEQC.

Bee, M. and Dolton, P. (1985) Degree class and pass rates: an inter-university comparison, *Higher Education Review*, 7, 45-52.

Bourner, T. and Hamed, M. (1987) *Entry qualifications and degree performance*, CNAA, London.

Bradley, C. (1984) Sex bias in the evaluation of students, *British Journal of Social Psychology*, **23**, 147-173.

Bradley, C. (1993) Sex bias in student assessment overlooked?, *Assessment and evaluation*, **18**, 3-8.

Browne, W.J., Goldstein,H. and Rasbash, J. (2001) Multiple membership multiple classification (MMMC) models, *Statistical Modelling*, to appear.

Bridges,P. Bourdillon, B. Collymore,D., Cooper, A., Fox, W., Haines, C. Turner,D. Woolf,H. Yorke,M. (1999) Discipline-related marking behaviour using percentages: a potential cause of inequity in assessment, *Assessment and Evaluation in Higher Education*, **24**, 285-300.

Bristow, S. (1989) Resource changes and academic performacne at an English polytechnic - a comment to Lindsay and Paton-Salzberg, *Studies in Higher Education*, **14**, 103-104.

Brown, S. and Knight, P. (1994) *Assessing Learners in Higher Education*, London, Kogan Page.

Browne, W.J. (1998) *Applying MCMC Methods to Multilevel Models, PhD dissertation, Department of Mathematical Sciences*, University of Bath, UK.

Browne, W.J. (2000) Software, personal communication.

Browne, W.J. and Draper, D. (2000) Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. To appear in *Computational Statistics.*

Browne, W.J., Draper, D., Goldstein, H. and Rasbash, J. (2000) A comparison of methods for fitting multilevel models with complex level 1 variation. To be submitted.

Carlin, B.P. and Louis, T.A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall, London.

Chapman, K. (1996) Entry qualifications, degree results and value-added in UK universities, *Oxford Review of Education*, **22**, 251-264.

Clarke, S. (1988) Another look at the degree results of men and women, *Studies in Higher Education*, **13**, 315-331.

Clayton, D. and Rasbash, J. (1999) Estimation in large crossed random effect models by data augmentation, *Journal of the Royal Statistical Society, A*, **162**, 425-436.

Cohen, G. and Fraser, E.J.P. (1992) Female participation in mathematical degrees at English and Scottish universities, *Journal of the Royal Statistical Society, A*, **155**, 241-258.

Cresswell, M. (1988) Combining grades from different assessments: how reliable is the result?, *Educational Review*, **40**, 361-382.

Cresswell, M. (1996) Defining, setting and maintaining standards in curriculum-embedded examinations, in : H. Goldstein and T. Lewis , *Assessment: problems, developments and statistical issues*, Chichester, Wiley.

Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom (CVCP) (1984) *The External Examiner System for First Degree and Taught Master's Courses*, CVCP, London.

Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom (CVCP) (1995) *Briefing note: The growth in student numbers in British higher education*, CVCP,London.

Dennis, I. And Newstead, S.E.(1994) The strange case of the disappearing sex bias, *Assessment and Evaluation in Higher Education*, **19**, 49-56.

Elton, L. (1998) Are UK degree standards going up, down or sideways?, *Studies in Higher Education*, **23**, 35-42.

Entwhistle, N. and Ramsden, P. (1983) *Understanding student learning*, London, Croom Helm.

Fearnley, S. (1995) Class size: the erosive effect of recruitment numbers on performance, *Quality in Higher Education*, **1**, 59-65.

Ford, J., Bosworth, D. and Wilson, R. (1995) Part-time work and full-time higher education, *Studies in Higher Education*, **20**, 187-202.

Franklyn-Stokes, A. and Newstead, S.E. (1995) Undergraduate cheating: who does what and why?, *Studies in Higher Education*, **20**, 159-172.

Gibbs, G. and Lucas, L. (1997) Coursework assessment, class size and student performance, *Journal of Further and Higher Education*, **21**, 183-192.

Gibbs, G., Lucas, L. and Simonite, V. (1996) Class size and student performance:1984-194, *Studies in Higher Education*, **21**, 261-273.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Gipps,C. and Murphy, P. (1994) *A Fair Test? Assessment, Achievement and Equity*, Buckingham, Open University.

Goldstein, H. (1995) *Multilevel statistical models*, second edition, Arnold,London.

Goldstein,H. and Cresswell,M. (1996) The comparability of different subjects in public examinations: a theoretical and practical critique, *Oxford Review of Education*, 22, 435-442..

Good, F. and Cresswell, M. (1988) *Grading the GCSE*, London, The Secondary examinations Council.

Hartley, J., Trueman, M. and Lapping,C.,(1997) The performance of mature and younger students at Keele University: an analysis of archival data, *Journal of Access Studies*, **12**, p98-112.

Hayes, K. and Richardson, J.T.E (1995) Gender, subject and context as determinants of approaches to studying in higher education, *Studies in Higher Education*, **20**,No 2, 215-221.

Higher Education Quality Council (1996) *Modular Higher Education in the UK: in focus*, London, HEQC.

HEQC (1996a) *Inter-Institutional variation in degree results*, London, HEQC.

HEQC (1996b) *Understanding academic standards in modular frameworks*, London, HEQC.

HEQC (1997) *Assessment in higher education and the role of graduateness*, London, HEQC.

Higher Education Statistics Agency (1997*) Students in Higher Education Institutions 1995/6*, Cheltenham, HESA.

Higher Education Statistics Agency (1998) *Students in Higher Education Institutions 1996/7*, Cheltenham, HESA.

Hoskins, S.L., Newstead, S.E. and Dennis, I. (1997) Degree performance as a function of age, gender, prior qualifications and discipline studied, *Assessment and Evaluation in Higher Education*, **22**, 317-328.

Jackson, N. (1996) Assessment issues, in HEQC (1996) *Modular Higher Education in the UK: in focus*, London, HEQC.

Johnes, J. (1992) The potential effects of wider access to higher education on degree quality, *Higher Education Quarterly*, **46**, 88-107.

Johnes, J. and Taylor, J. (1987) Degree quality: an investigation into differences between UK universities, *Higher Education* **16**, 581-602.

Johnson, S. (1988) Comparability in degree awards: implications of two decades of secondary level examinations research, *Studies in Higher Educa*tion, **13**, 177-187.

Kniveton, B.H (1996) Student Perceptions of Assessment Methods, *Assessment and Evaluation in Higher Education*, **21**, 229-237.

Leopold,J.W. and Osborne,M. (1996) The performance of former in-house Access students at a Scottish university, *Journal of Access S*tudies, **11**, 120-131.

Lindsay,R.O. and Paton-Salzburg,R. (1987) Resource changes and academic performance at an English polytechnic, *Studies in Higher Education*, **12**, 213-227.

MacFarlane,B. (1998) Degree classifications: time to bite the bullet, *Teaching in Higher Education*, **3**, 401-405.

MacFarlane,B. (1992) The results of recession: students and university degree performance during the 80s, *Research in Education*, **49**, 1-49.

Mahler,S. , Neumann, L. and Tamir, P. (1986) The class-size effect upon activity and cognitive dimensions of lessons in higher education, *Assessment and Evaluation in Higher Education*, **11**, 43-59.

Metcalf, H. (1993) *Non-traditional students' experience of higher education*, London, CVCP.

Meyer, J.H.F. (1995) Gender-group differences in the learning behaviour of entering first-year university students, *Higher Education*, **29**, 201-215.

Modular Degree Programme (1994) *Undergraduate Modular Programme: Handbook and Regulations*, Oxford, Oxford Brookes University.

Morrison, H, Cowan, P. and Harte, S. (1997) The impact of modular aggregation on the reliability of final degrees and the transparency of European Credit Transfer, *Assessment and Evaluation in Higher Education*, **22**, 405-417.

National Committee of Inquiry into Higher Education (1997) *Higher education in the learning society, Report of the National Committee (The Dearing Report)*, London, HMSO.

Newstead, S.E. and Dennis, I. (1990) Blind marking and sex bias in student assessment, *Assessment and Evaluation in Higher Education*, **15**, 132-139.

Newton, P.E. (1997) Measuring comparability of standards between subjects: why our statistical techniques do not make the grade, *British Educational Research Journal*, 23, 433-449.

Nuttall, D.L. (1979) The myth of comparability, in Murphy,R. and Broadfoot,P. (1995) *Effective Assessment and the Improvement of Education: a tribute to D. Nuttall*, Falmer, London.

Paton-Salzberg, R. and Lindsay, R.O. (1996) *The Effects of Paid Employment on the Academic Performance of Full-time Students in Higher Education: A report of a study commissioned by the Academic Standards Committee of Oxford Brookes University*, Oxford, Oxford Brookes University.

Peers, I.S. (1994) Gender and age bias in the predictor-criterion relationship of A levels and degree performance: a logistic regression analysis, *Research in Education*, 52, 23-41.

Peers, I.S. and Johnston, M. (1994) Influence of Learning Context on the relationship between A-levels Attainment and Final Degree Performance, *British Journal of Educational Psychology*, **64**, 1-18.

Raftery,A.E. and Lewis,S.M. (1992) How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (eds. J.M. Bernado, J.O.Berger, A.P.Dawid and A.F.M. Smith), 675-76. Oxford , Oxford University Press.

Rasbash, J., Browne, W., Goldstein, H.,Yang, M. Plewis, I.,Healy, M. Woodhouse, G., Draper, D. Langford, I.and Lewis, T. (2000) *A User's Guide to MlwiN, second edition*,London, Institute of Education.

Rasbash, J. and Goldstein,H.(1994) Efficient analysis of mixed Hierarchical and cross-classified random structures using a multilevel model, *Journal of Educational and Behavioural Statistics*, **19**, 337-350.

Richardson,J.T.E. (1993) Gender differences in responses to the apporaches to studying inventory, *Studies in Higher Education*,**18**, 3-13.

Richardson,J.T.E. (1994) Mature students in higher education: academic performance and intellectual ability, *Higher Education*, **28**, 373-486.

Richardson,J.T.E. (1995) Mature students in higher education : II.An investigation of approaches to studying and academic performance, *Studies in Higher Education*, **20**, 5-17.

Rudd, E. (1984) A comparison between the results achieved by women and men studying for first degrees in British universities, *Studies in Higher Education*, 9, 47-57.

Rudd,E. (1987) Students and social class, *Studies in Higher Education*, **12, 99-106.**

Schafer,J.L. (1997) *Analysis of Incomplete Multivariate Data*, Chapman and Hall, London.

Sear,K. (1983) The correlation between 'A' level grades and degree results in England and Wales, *Higher Education*, **12**, 609-619.

Silver, H., Stennett,A. and Williams,R. (1995) *The external examiner system: possible futures*, HEQC, London.

Simonite, V. (1997) Academic achievement of mature students on a modular degree course, *Journal of Further and Higher Education*, **21**, 241-249.

Somerville, J. (1996) The logic of modularity: the challenge to academic authority, in: HEQC (1996) *Modular Higher Education in the United Kingdom: in focus*, pp.27-32, London, HEQC.

Tanner, M.A. and Wong W.H. (1987) The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, **82**, 528-550.

Tomlinson, K. and MacFarlane, B. (1995) The significance of subject choice in explaining the first-class degree divide between male and female graduates, *Research in Education*, **54**, 95-101.

Walker,P. (1975) The university performance of mature students, *Research in Education*, **14**,1-13.

Warren Piper, D. (1986) Unpublished paper, quoted in HEQC (1996) *Modular Higher Education in the United Kingdom: in focus*, London, HEQC.

Watson, D. (1989) *Managing the Modular Course*, SGRE/OU Press, Milton Keynes.

Winter,R. (1993) Education or grading? Arguments for a non-subdivided honours degree, *Studies in Higher Education*, **18**, 363-377.

Wood, R. (1991) *Assessment and Testing: a survey of research*, Cambridge, Cambridge University Press.

Woodley,A. (1984) The older the better? A study of mature student performance in British universities, *Research in Education*, **32**, 35-50.

Yorke,M., Cooper,A.,Haines,C.,Turner,D.,Fox,W.,McHugh,P. and Woolf,H. (1996) Module mark distributions in eight subject areas and some issues they raise, in HEQC (1996) *Modular Higher Education in the UK: in focus*, HEQC, London.